Neural dynamics of selective attention to speech in noise

Neural dynamics of selective attention to speech in noise

Der Fakultät für Biowissenschaften, Pharmazie und Psychologie

der Universität Leipzig

eingereichte

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von

Malte Wöstmann, M.Sc. (Cognitive Science)

geboren am 02.04.1986 in Mettingen

Leipzig, 16.04.2015

Die Arbeit wurde erfolgreich verteidigt am 08.10.2015

Acknowledgements

I would like to thank several people who supported my work within the last 2 ½ years at the Max Planck Institute to make this thesis possible.

First and foremost, I would like to thank my supervisor Jonas Obleser for teaching me how to conduct high-quality science, for supporting me in turning my research interests into experiments, for his great assistance in analysing data and discussing results, for reinforcing me to present my research in public talks and journal articles, and for substantially promoting my career in science.

I would like to thank the co-supervisors of my research, Erich Schröger, Burkhard Maess, and Thomas Pechmann. Scientific exchange with you about methodological questions and theoretical considerations improved this thesis substantially. Our discussions in the annual thesis committee meetings helped me to stay focused at all stages of this thesis.

I am grateful that I could be part of the research group "Auditory Cognition" at the Max Planck Institute in Leipzig for the last 2 ½ years. I am obliged to Dunja Kunke, Mathias Scharinger, Molly Henry, Alex Brandmeyer, Julia Erb, Antje Strauß, Anna Wilsch, Sung-Joo Lim, Sophie Herbst, and Lorenz Fiedler for fruitful discussions in Tuesday's group meetings and after-lunch coffee breaks, for two great retreats with you in Bad Sachsa and Königs Wusterhausen, and for after-work beers with you at Café Cantona. In particular, I am grateful for Björn Herrmann's support in data analyses and methodological considerations of my research. Furthermore, I would like to thank all present and former student assistants of the "Auditory Cognition" group for their help in recruiting participants and collecting data. Particular thank goes to Steven Kalinke who assured good quality of the layout of this thesis.

I would like to thank collaborators at the Eriskholm research centre in Denmark, Eline Borch-Petersen, Thomas Lunner, and Niels Henrik Pontoppidan. I very much enjoyed my two visits at the Eriksholm research centre in 2012 and 2015 and learned a lot from you.

I thank my family, friends, and especially Sina for their unconditional support and sympathy.

Malte Wöstmann, Leipzig, April 2015

Bibliographic details

Malte Wöstmann, *Neural dynamics of selective attention to speech in noise*

Fakultät für Biowissenschaften, Pharmazie und Psychologie der Universität Leipzig

Dissertation

177 pages, 361 references, 28 figures

This thesis investigates how the neural system instantiates selective attention to speech in challenging acoustic conditions, such as spectral degradation and the presence of background noise. Four studies using behavioural measures, magneto- and electroencephalography (M/EEG) recordings were conducted in younger (20–30 years) and older participants (60–80 years). The overall results can be summarized as follows. An EEG experiment demonstrated that slow negative potentials reflect participants' enhanced allocation of attention when they are faced with more degraded acoustics. This basic mechanism of attention allocation was preserved at an older age. A follow-up experiment in younger listeners indicated that attention allocation can be further enhanced in a context of increased task-relevance through monetary incentives. A subsequent study focused on brain oscillatory dynamics in a demanding speech comprehension task. The power of neural alpha oscillations (~10 Hz) reflected a decrease in demands on attention with increasing acoustic detail and critically also with increasing predictiveness of the upcoming speech content. Older listeners' behavioural responses and alpha power dynamics were stronger affected by acoustic detail compared with younger listeners, indicating that selective attention at an older age is particularly dependent on the sensory input signal. An additional analysis of listeners' neural phase-locking to the temporal envelopes of attended speech and unattended background speech revealed that younger and older listeners show a similar segregation of attended and unattended speech on a neural level. A dichotic listening experiment in the MEG aimed at investigating *how* neural alpha oscillations support selective attention to speech. Lateralized alpha power modulations in parietal and auditory cortex regions predicted listeners' focus of attention (i.e., left vs right). This suggests that alpha oscillations implement an attentional filter mechanism to enhance the signal and to suppress noise. A final behavioural study asked whether acoustic and semantic aspects of task-irrelevant speech determine how much it interferes with attention to task-relevant speech. Results demonstrated that younger and older adults were more distracted when acoustic detail of irrelevant speech was enhanced, whereas predictiveness of irrelevant speech had no effect. All findings of this thesis are integrated in an initial framework for the role of attention for speech comprehension under demanding acoustic conditions.

Graphical summary. The wordcloud shows a random arrangement of the most frequent content words used in this thesis. Larger fontsizes indicate more frequent occurrences of individual words.

Contents

1 General introduction

Human environments are rich of sensory information that compete for limited cognitive processing capacities (for review, see Marois and Ivanoff, 2005). Selective attention describes the mental faculty of selecting currently relevant information for further processing at the expense of irrelevant distractors (Desimone and Duncan, 1995). Human speech is a paradigmatic case of a sensory signal that notoriously occurs in the presence of irrelevant information originating from environmental noise or competing talkers (e.g., Pichora-Fuller et al., 1995). The success of selective auditory attention depends on a listener's individual attentional capacity but also on hearing acuity (Shinn-Cunningham and Best, 2008). In the present thesis, I will report four studies with younger and older human participants to investigate the neural dynamics of speech comprehension in attention-demanding listening conditions. The goal of this research is to foster our understanding of the neural mechanisms that support selective attention and might thus compensate for degradations of the sensory input.

1.1 Selective attention

Unless stated otherwise, the present thesis is concerned with volitional (i.e., *top-down*) attention to task-relevant auditory signals. Moreover, this thesis explores the neural dynamics of *selective attention* to one auditory stimulus in an environment of potential distractors. To the contrary, this thesis does not investigate the automatic (i.e., *bottom-up*) capture of attention by unexpected and salient stimuli or *divided attention*, which describes the mechanisms of dividing and switching of attention between multiple stimuli.

Research on selective attention is traditionally concerned with the question at which stage of information processing the cognitive system "filters out" unattended stimuli, known as the *early vs late selection debate* (Serences and Kastner, 2014). Although the question of early vs late selection is not the major focus of this thesis, I will briefly outline these two opposing accounts and elucidate on possible implications for speech processing.

1.1.1 Early vs late selection

The *early selection* account holds that sensory information is processed up to a level where basic physical features (e.g., pitch, location) are analysed. Based on these basic features, irrelevant information can be identified and filtered out to hinder it from further processing (Broadbent, 1958).

Importantly, this implies that irrelevant stimuli are filtered out before higher-level features such as semantic information are extracted. Experimental support for the early selection account came from dichotic listening studies, where listeners were presented with two simultaneous speech streams; one on the left and the other on the right ear (see also Study 3; for a review of the cognitive and neural bases of dichotic listening, see Hugdahl et al., 2009). When listeners were attending and verbally repeating (i.e., *shadowing*) speech from one side, they were able to report basic physical features such as the gender of the unattended speaker but they were unable to report the semantic content or individual words of the unattended speech (e.g., Cherry, 1953; Moray, 1959). This suggests that the unattended signal is filtered out on the level of basic feature analysis and is not processed further (for a more recent series of experiments in support of early selection, see Lachter et al., 2004).

To the contrary, the *late selection* account asserts that attended and unattended signals are processed automatically and in parallel up to a level where semantic information is analysed (Deutsch and Deutsch, 1963). Selective attention operates on this sematic information to filter out the unattended signal. The late selection account received initial empirical support from the finding that some listeners detected their own names in the unattended stream during dichotic listening (Moray, 1959). Later, it was found that individual differences in working memory affected the probability of listeners' detection of their own names in unattended speech (Conway et al., 2001; Colflesh and Conway, 2007). These findings suggest that parts of unattended speech are processed up to a level of semantic analysis.

Taking into account the empirical evidence for both early and late selection, Anne Treisman proposed an *attenuation model* of selective attention that implemented early and late selection depending on the level of internal thresholds for stimuli (Treisman, 1960, 1964). This model is a modification of Broadbent's early selection model in so far as unattended stimuli are not filtered out but are rather attenuated on the basis of their basic physical features. After attenuation, each stimulus is compared against its own threshold and analysed further if this threshold is exceeded. For instance, a listener's own name has a low threshold, which could explain why it is often detected even if it is attenuated since it is part of an unattended speech stream.

More recent theories of selective attention suppose that the level of attentional selection (early vs late) is not fixed but depends on task demands and perceptual load (e.g., Huang-Pollock et al., 2002; Lavie et al., 2004; Yi et al., 2004; Cartwright-Finch and Lavie, 2007). Nilli Lavie (2005, 2010) proposed that high perceptual load (i.e., a large number of attended sensory stimuli) decreases the degree of

processing unattended stimuli, thereby implementing early selection. In contrast, high cognitive load (e.g., high working memory demand) increases processing of unattended stimuli, which is compatible with a late selection account of attention.

1.1.2 What is special about selective attention in the auditory modality?

A common assumption of all models of selective attention described above is that attended and unattended signals are divided into distinct "objects" on the basis of their physical features. Attention selects some of these objects for further processing while others are filtered out. However, what happens if the basic physical features of two signals overlap, such as in the case of two female speakers with similar pitch, loudness, and spatial location? Obviously, selective attention depends on the accurate definition of objects.

In analogy to the theory of object-based attention in the visual modality (for reviews, see Scholl, 2001; Chen, 2012), Barbara Shinn-Cunningham (2008) proposed that the concept of perceptual objects holds also for auditory selective attention. Simply speaking, an auditory object describes a collection of sounds emitted from one physical source. In a busy cafeteria, individual talkers, clinking glasses, and the sounds of the checkout counter would constitute separate auditory objects. Object-based attention asserts that attention operates on objects rather than individual features (Kubovy and Van Valkenburg, 2001). Thus, attention to one feature of an object also increases sensitivity to other features of the same object (in the visual modality: Duncan, 1984; across modalities: Busse et al., 2005; in the auditory modality: Best et al., 2008).

But is there a fundamental difference in the formation of visual and auditory objects? Since visual objects are often spatially distinct, it has been proposed that saliency maps guide visual attention (e.g., Itti and Koch, 2000). Furthermore, humans can shift their direction of gaze to enhance the perceptual encoding of relevant visual objects in the fovea (Treue, 2003). On the contrary, auditory object formation is particularly vulnerable to distractor interference. Auditory stimuli at each point in time can be characterized along several dimensions such as pitch, intensity, and location. In the human cochlea, however, auditory stimuli are initially represented on a one-dimensional spectral axis (Shamma, 2001). Since human speech has a rich spectro-temporal structure, two simultaneously presented speech signals likely evoke overlapping excitation patterns in the cochlea (Moore, 2008a). Thus, it has been proposed that temporal coherence promotes object formation in the auditory modality (Shamma et al., 2011). In the most extreme case, challenging listening conditions (see below)

can render signal and distractor inseparable, resulting in a failure of auditory object formation and deficient selective attention.

If auditory object formation is successful, currently relevant objects have to be selected for further processing. Object selection is a top-down process, such that a listener selects an object at will, for instance because it is relevant for the current task. Object selection requires attentional control (Hopfinger et al., 2000), i.e., the prioritization of relevant objects for further processing. Object selection is compromised if listeners are uncertain about which auditory object to select (e.g., Kidd et al., 2005) or if irrelevant auditory objects are highly salient (e.g., Passow et al., 2012). In sum, selective attention to speech in complex acoustic environments critically depends on the successful formation and selection of auditory objects.

1.1.3 Selective attention in older and hearing-impaired listeners[1]

In acoustically demanding multi-talker situations, older listeners typically experience more difficulties compared with younger adults. It is however unclear, in how far these difficulties are caused by age-related decline in perceptual auditory acuity (hearing loss or loss of temporal and spectral resolution; Fostick and Babkoff, 2013), decline of cognitive functioning with age, or both (Wingfield et al., 2005). Critically, both auditory perceptual and cognitive decline could lead to insufficient selective attention. First, compared to normal-hearing controls, listeners with hearing loss are less successful in utilizing spectral (Lorenzi et al., 2006), temporal (Tremblay et al., 2003), and spatial auditory cues (Neher et al., 2009) important for the perceptual segregation of different sound sources (i.e., auditory object formation). Thus, attending to relevant and inhibiting irrelevant sound sources is impaired, as auditory features are lacking to distinguish the different sound sources in the first place (Shinn-Cunningham and Best, 2008). Second, age negatively affects many aspects of cognitive functioning (Park et al., 2003), amongst it the ability to suppress irrelevant but salient auditory distractors (Chao and Knight, 1997; Tun et al., 2002; Passow et al., 2014). Thus, even if the perceptual segregation of sound sources is accomplished successfully, the insufficient inhibition of distractors may constrain auditory object selection.

What is the practical significance of studying the underlying neural mechanisms of speech perception in older listeners? Changes in the neural dynamics of speech processing could serve as an

[1] This section is largely adapted from the article published in Frontiers in Human Neuroscience by Straußt, Wöstmannt, & Obleser (2014). † both authors contributed equally.

indicator of age-dependent decline in selective attention. Auditory selective attention might function as a compensatory mechanism as listening conditions become more demanding, for instance due to a decreasing signal-to-noise ratio (SNR). The study of neural dynamics could help to reveal how listeners of different age exert top-down attentional control to enhance processing of task-relevant signals and inhibit processing of interfering distractors. In particular, this line of research might foster the understanding of why older listeners find it more exhausting to participate in cocktail party-like listening situations compared to younger listeners (Pichora-Fuller, 2003b).

1.2 Neural bases of selective attention

To study the neural bases of attention, it is important to take into account some theoretical considerations. In general, it is necessary to differentiate two complementary mechanisms of attention; first, the enhancement of the signal, and second, the suppression of noise (e.g., Yeshurun and Carrasco, 1998; O'Connor et al., 2002; Gazzaley et al., 2005b). In detail, changes in brain activity during selective attention compared to a baseline could reflect (a) the enhancement of the attended signal, (b) the suppression of noise, (c) both, or (d) other brain processes not directly related to selective attention, such as conflict monitoring (e.g., Botvinick et al., 2004; Yeung et al., 2004). Importantly, suppression of noise does not exclusively refer to noise in the external stimulation but can as well refer to neural noise (e.g., Briggs et al., 2013), e.g., activity in task-irrelevant brain regions (e.g., Lawrence et al., 2003; Polk et al., 2008; Snyder and Foxe, 2010). For instance, it has been suggested that auditory attention involves an inhibition of activity in the visual system (e.g., Adrian, 1944; Fu et al., 2001; Johnson and Zatorre, 2005).

Similarly to the present thesis, research on selective attention is often not directly concerned with the question of signal enhancement or noise suppression. It is nevertheless useful to consider this distinction when it comes to the interpretation of neuroimaging results. If, however, the goal is to investigate signal enhancement and noise suppression independently, it is not sufficient to hold one of the two constant and to manipulate the other. For example, a change in brain activity when external noise is reduced could both reflect more thorough neural noise suppression but also more thorough neural signal enhancement. In order to explore these two mechanisms unambiguously, the neural responses to signal and noise have to be separated, which can be realized by spatial separation of signal and noise in dichotic listening tasks (e.g, Alho et al., 2012; see also Study 3) or by sequential presentation of signal and noise (e.g., Kastner et al., 1998; see also Study 4). The following section will

briefly summarize evidence for the structural and functional foundations of attention networks in the brain. Thereafter, I will outline the electrophysiological bases of selective attention to set the stage for the magneto-and electroencephalography (M/EEG) studies reported in this thesis.

1.2.1 Neural attention networks

Before the advent of the wide availability of functional magnetic resonance imaging (fMRI), neuroscientific research on selective attention primarily investigated activity changes within spatially restricted regions of the brain. Single cell recordings in monkeys revealed that attention increases the firing rate (e.g., Motter, 1993) and sharpens the tuning curves of neurons sensitive to attended stimuli (e.g., Spitzer et al., 1988; for review, see Treue, 2001). Later, modulation of sensitivity in primary sensory areas during attention was also evidenced through changes of the fMRI blood oxygen level dependent (BOLD) signal in visual (e.g., Brefczynski and DeYoe, 1999) and auditory cortex regions (e.g., Petkov et al., 2004). Critically, whole-brain fMRI opened up the possibility to explore whether also brain regions other than sensory areas underlie human selective attention.

Indeed, major evidence for the contribution of a complex network of brain regions to attention comes from fMRI investigations and brain lesion studies. A number of fMRI experiments found that frontal and parietal brain regions (including the frontal eye field, FEF, superior parietal lobe, SPL, and the intraparietal sulcus, IPS) show increased levels of activation in spatial attention tasks (e.g., Corbetta et al., 1998; Kastner et al., 1999; Kim et al., 1999), suggesting that these areas constitute an *attention network*. It has been proposed that this attention network can be further divided into smaller, more specialized sub-systems (e.g., Pourtois et al., 2006; Shallice et al., 2008). For instance, a network of intraparietal and superior frontal cortex guides goal-directed visual attentional selection, whereas a network of temporoparietal and inferior frontal cortex directs attention to unexpected and salient events (Corbetta and Shulman, 2002). Another line of research revealed that patients with brain lesions in diverse brain regions exhibit an impairment of spatial attention, i.e., spatial hemineglect (Vallar, 1998). Spatial hemineglect was found to correlate with lesions in the parietal lobe (e.g., Vallar and Perani, 1987), but also in the frontal (e.g., Damasio et al., 1980) and temporal lobe (Karnath et al., 2001), supporting the contribution of widespread brain regions to the attention network. Thus, converging evidence from fMRI and brain lesion studies suggests that selective attention is instantiated within a distributed network of sensory, parietal, and frontal brain regions (for review, see Pessoa et al., 2003).

But in how far is communication between distinct areas within the attention network essential to establish attention? To test this, it is necessary to analyse the functional connectivity among brain regions, involving methods of graph theory (Bullmore and Sporns, 2009). In a recent combined fMRI/MEG study, Baldauf and Desimone (2014) found that during attention to faces and houses, the inferior frontal junction exhibited functional connectivity (via gamma oscillations) with the fusiform face area (FFA) and the parahippocampal place area (PPA), respectively. This suggests that functional interactions between frontal and sensory areas underlie the control of selective attention. Critically, results from transcranial magnetic stimulation (TMS) studies revealed that activity in frontal cortex modulates activity in sensory regions during attention (Smith et al., 2009; Zanto et al., 2011a), suggesting that frontal regions exert top-down control to regulate sensitivity in sensory areas. In sum, the neural bases of attention comprise a network of frontal and parietal cortex regions that functionally connect with primary sensory regions (for review, see Ptak, 2012) to enhance relevant and to suppress irrelevant sensory information.

1.2.2 Electrophysiological bases of selective attention

One of the most well-known effects of attention in electroencephalography (EEG) studies is the increased amplitude of early event-related potential (ERP) components (Hillyard et al., 1973; Näätänen et al., 1978; for review, see Luck et al., 2000). Compared to a passive control condition, evoked responses increase when participants attend to visual (e.g., Heinze et al., 1990) and auditory stimuli (e.g., Woldorff et al., 1987). Larger ERP amplitude has been interpreted as an increase in attention orientation and facilitation of sensory processing (e.g., Luck et al., 1990). Interestingly, enhanced prediction of stimuli shows the opposite effect, i.e., a reduction of the neural response (e.g., Arnal and Giraud, 2012; Kok et al., 2012). This suggests that attention is not a self-contained neural mechanism but interacts with predictions to guide perception (Summerfield and Egner, 2009; Schröger et al., 2015).

Attention also affects the ERP prior to stimulus onset (e.g., Walter et al., 1964; Weinberg, 1972). Slow cortical potentials increase during the anticipation of upcoming stimuli (for review, see Van Boxtel and Böcker, 2004). If distraction impairs attention, the magnitude of slow cortical potentials decreases (e.g., McCallum and Walter, 1968; Tecce and Scheff, 1969). To the contrary, larger potential magnitude improves stimulus detection (e.g., Rockstroh et al., 1993; O'Connell et al., 2009), indicating increased selective attention. Considering the functional role of slow cortical potentials, it has been suggested that larger magnitudes reflect the attentional enhancement of sensitivity in task-relevant

cortical networks (Raichle, 2011). Important for the present thesis, slow cortical potentials are useful to study how situational factors such as degraded acoustics (Study 1.1) or higher levels of motivation (Study 1.2) affect the allocation of attention to ensuing stimuli (e.g., Rebert et al., 1967).

Another electrophysiological correlate of selective attention that is particularly relevant to the study of speech stimuli with a rich temporal structure is the alignment (i.e., *phase-locking*) of the ongoing M/EEG signal to the temporal structure of stimuli (e.g., Horton et al., 2013; O'Sullivan et al., 2014). While the M/EEG signal aligns with the temporal envelope of attended speech, it shows a characteristically distinct alignment with concurrent unattended speech (e.g., Ding and Simon, 2012; Mesgarani and Chang, 2012; Zion-Golumbic and Schroeder, 2012; see also Study 2.2). This mechanism might instantiate the co-occurrence of phases of high neural excitability with most critical segments of attended speech and phases of low excitability with unattended speech (Giraud and Poeppel, 2012). Thus, differential alignment of neural responses with attended and unattended speech might constitute a neural mechanism to select task-relevant speech for higher order processing (Zion Golumbic et al., 2013).

Besides stimulus-locked neural activity (see above), neural oscillations which are not strictly time-locked to the stimulus reflect attentional processes (for review, see Herrmann and Knight, 2001). More than 70 years ago, Edward Douglas Adrian (1944) observed that the amplitude of parietal alpha oscillations with a frequency of approximately 10 Hz increases when participants direct attention to auditory stimuli. Considering that alpha amplitude also increases if participants close their eyes, Adrian supposed that high alpha amplitude indicates "inattention" of the visual system, which is "unemployed" during auditory attention. In agreement with Adrian's observations, a large number of more recent studies suggest that alpha activity reflects the inhibition of task-irrelevant neural processes to support attention to relevant stimuli (e.g., Klimesch et al., 1999; Jensen et al., 2002; for reviews, see Jensen and Mazaheri, 2010; Foxe and Snyder, 2011). The inhibitory effect of alpha activity is further confirmed by a negative correlation with the fMRI BOLD signal (e.g., Laufs et al., 2003; Scheeringa et al., 2009; Yuan et al., 2010). Thus, alpha activity is a possible neural mechanism of noise suppression to support attention (see also Studies 2.1 & 3). Although alpha oscillations dominate measures of the human M/EEG, it is important to note that they co-occur and interact with oscillations in other frequency bands (e.g., Lakatos et al., 2005; Spaak et al., 2012; Roux and Uhlhaas, 2014). For instance, alpha oscillations are coupled to gamma oscillations (>30 Hz), which are thought to be involved in active neural processing (e.g., Palva et al., 2005; Osipova et al., 2008). Particularly relevant for the

present thesis, alpha oscillations reliably reflect attentional demands in speech comprehension tasks (e.g., Obleser et al., 2012; Becker et al., 2013; Meyer et al., 2013).

In summary, electrophysiological measures of neural activity provide several signatures of attention (ERP amplitude, neural phase-locking, alpha oscillations), which are extracted from the same underlying M/EEG signal. The following section will outline in how far the study of these neural signatures of attention might answer the research questions of the present thesis.

1.3 Research questions

The present thesis comprises four studies which investigate behavioural responses and electrophysiological recordings of neural activity in an overall sample of 98 participants. This thesis aims at exploring the neural mechanisms that support selective attention to speech under spectral deterioration and the presence of background noise, i.e., acoustic degradation. A further objective is to understand the vulnerability of attention mechanisms in populations which experience particular difficulties of selective attention. Thus, three studies of this thesis compare healthy older adults (60–80 years) to younger adults (20–30 years).

This thesis tries to answer four major research questions: (1) Do acoustic conditions of the external stimulation guide listeners' allocation of attention in demanding listening situations? Study 1 investigates whether ERP signatures of attention reflect an increased allocation of attention when listeners are faced with more degraded acoustic input. To test the susceptibility of attention allocation in listeners with difficulties in complex acoustic situations, it is further explored whether neural dynamics of attention allocation are preserved at an older age. A follow-up experiment with younger participants (Study 1.2) examines in how far ERP signatures of attention are affected by task-relevance manipulated by monetary incentives. Subsequently, this thesis explores whether in addition to external acoustic conditions, also listeners' formation of predictions about the upcoming speech content affect the neural dynamics of selective attention. Study 2 asks: (2) Do increased acoustic detail on the one hand and better predictions of the speech content on the other hand facilitate neural mechanisms of selective attention to the same degree? Study 2.1 investigates how these two factors affect the power of neural alpha oscillations; i.e., a neural signature that is thought to reflect the inhibition of task-irrelevant brain processes. Behavioural responses and alpha power modulations in younger and older listeners are compared to test to what extent attentional control changes at an older age. In an

additional analysis, Study 2.2 tests whether age-differences in attentional control do also show up in the neural phase-locking to attended and unattended speech signals.

After demonstrating that neural alpha oscillations are sensitive to demands on selective attention to speech in noise, this thesis turns to the question *how* alpha power dynamics support attention. Thus, Study 3 asks: (3) In how far do alpha power modulations implement an *attentional filter* to suppress noise and to enhance the task-relevant speech signal? To test this, Study 3 uses a dichotic listening task in the MEG to investigate whether modulations of neural alpha power in a network of parietal and auditory cortex regions reflect the attentional selection of speech in noise. In particular, this study tests whether the deployment of spatial selective attention adapts to the temporal structure of ongoing speech. After investigating the electrophysiological bases of selective attention to speech in noise in Studies 1-3, Study 4 examines to what extent acoustic and semantic features determine the attentional capture of distracting speech (i.e., noise): (4) Do acoustic detail and predictability of a distracting speech signal determine how effectively it can be filtered out by selective attention? To test whether mechanisms of noise suppression change at an older age, this behavioural experiment again compares younger and older adults. In summary, this thesis aims at establishing an initial framework (see section 7.6) to explain the neural dynamics that support selective attention to speech in noise and thus counteract acoustic degradation.

2 Methodological background

This chapter will give a brief and general overview of the speech materials and different types of speech degradation used in the present thesis. Furthermore, I will introduce the neurophysiological basis of magneto- and electroencephalography (M/EEG) and different methods used for the data analysis. More detailed information can be found in the methods sections of Studies 1–4.

2.1 Stimulus materials: spoken digits

Stimulus materials used in the laboratory should generally fulfil two criteria. On the one hand, they should be of high internal validity, meaning that stimuli are well-controlled to exclude an impact of confounding factors on the results. On the other hand, stimuli should be of high external validity, meaning that they are representative of the studied phenomenon as it occurs in everyday life situations outside the laboratory. While it is usually difficult to meet both of these criteria, I will argue here that spoken digits are of high internal and external validity for the study of the neural dynamics of speech processing.

First, considering acoustic properties, spoken digits have a very regular temporal structure. For instance, German digits between 21 and 99 (used in Studies 1–3) all consist of four syllables uttered within approximately one second. Thus, spoken digits have a consistent length and a stable ~4 Hz syllable rate. Since M/EEG recordings have a high temporal resolution (see below), inconsistencies in the temporal structure of stimuli could increase the variance of neural activity measures. In contrast, the use of spoken digits as stimuli reduces measurement variability.

Second, considering the semantics of spoken digits, listeners have to understand several syllables of a single digit to process the numeric value correctly. For instance, in order to understand the German digit "ein-und-vier-zig" (*one-and-four-ty*), listeners have to understand at least the first and the third syllable. Thus, spoken digits possess an important property of natural speech, namely that information unfolds in time and that single units of information (i.e., syllables) have to be combined in order to derive meaning.

Third, all digits can be placed on the number line (Dehaene et al., 1998), which means that they are interrelated to each other: For each pair of digits, a listener can be asked whether the second digit was smaller or larger than the first (Studies 1&2). This is a simple question if the two digits are understood

correctly. However, what happens if one of the two digits is missed or only partly understood due to presence of distracting background noise? In such a case, listeners can use partial information to derive the most probable relationship between digits. For instance, if the first digit was very small, listeners might infer that the second digit was likely larger. Thus, spoken digits can be used to study an important cognitive mechanism for speech understanding in complex listening situations, i.e., the use of prior knowledge to infer the most likely meaning of an utterance (Study 2.1).

2.2 Speech degradation

Listening conditions in everyday life are rarely ideal. Speech can be degraded in several ways, ranging from poor acoustics on phone lines to complex multi-talker situations. This section will introduce two methods of speech degradation used in the present thesis, i.e., masking of speech by acoustic distractors (i.e., noise) and spectral degradation through vocoding.

2.2.1 Masking of speech by noise[2]

A characteristic feature of human sensory environments is that they are rich of information from relevant and irrelevant sources. Relevant and irrelevant signals are commonly referred to as target signals and noise/maskers, respectively. Whether a sensory signal is target or noise is often defined by the current goals and intentions of the perceiver. For example, a captivating audiobook can turn from a target signal into noise if the listener suddenly intends to listen to the lottery numbers announced on TV. Human speech is a paradigmatic case of a target signal that is often masked by different types of noises, which will be shortly described in the following.

Generally, it can be differentiated between energetic and informational masking. Energetic masking describes the competition of auditory target and masker in the auditory periphery due to spectro-temporal overlay of the two signals, causing an overlap of excitation patterns in the cochlea and auditory nerve (Durlach et al., 2003). One type of background signal often assumed to cause primarily energetic masking is white noise (e.g., Arbogast et al., 2005) which is quasi-stationary and has high energy in a broad frequency range (for discussion see Stone et al., 2012). Although informational masking is sometimes defined only negatively as all masking effects not accounted for by energetic masking (cf. Gutschalk et al., 2008), a more refined definition is required, especially when it comes to speech processing. When target speech is masked by a competing talker, it is not just the energetic

[2] This section is partly adapted from the article published in Frontiers in Human Neuroscience by Straußt, Wöstmannt, & Obleser (2014). † both authors contributed equally.

overlap of the two signals that causes interference. Rather, the speech masker initiates phonetic and semantic processing that interferes with the linguistic processing of the target (Schneider et al., 2007). Thus, informational masking describes the interference of target and masker at a more central, cognitive level, whereas energetic masking refers to energetic overlap in the auditory periphery. In this thesis, noise was primarily implemented by competing talkers, which cause energetic as well as informational masking.

2.2.2 Spectral degradation through vocoding

Like any other sound, human speech sounds are pressure waves propagated through air. In the human inner ear, these pressure waves cause vibrations of the basilar membrane in the cochlea. Hair cells transform these mechanical signals into electrical signals, which are then propagated to the brain via the auditory nerve. Since different frequencies are represented at different places on the basilar membrane, the inner ear can be modelled as a bank of filters that decomposes complex sounds into different frequency bands (Dau et al., 1997). Below, I describe a similar processing scheme as implemented in the human auditory system that is used as a basis for spectral degradation of speech through vocoding.

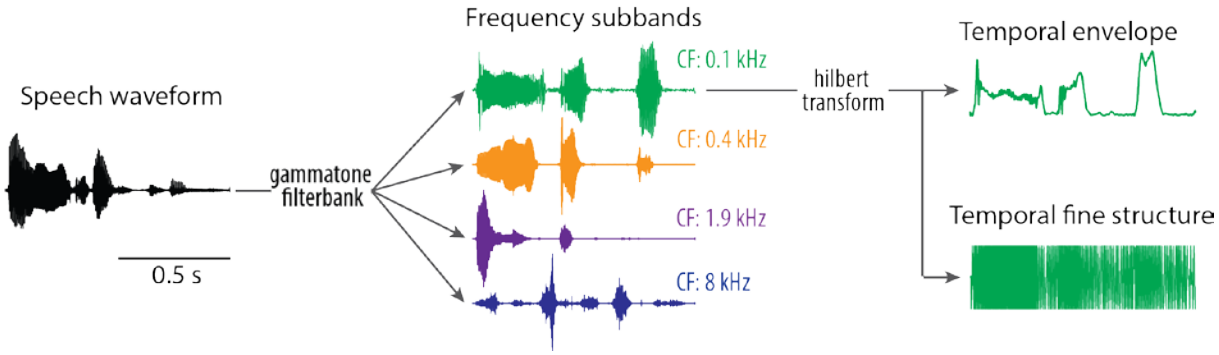

Figure 2.1. Decomposition of a speech sound into temporal and spectral information. Left: speech waveform of a German spoken digit ("61"). Middle: frequency subbands of the speech signal were derived by applying a bank of bandpass filters with logarithmically spaced center frequencies (CFs; 0.1, 0.4, 1.9, 8 kHz) to the speech signal (using a gammatone filterbank). Right: decomposition of one frequency subband into its slow temporal content (i.e., temporal envelope) and spectral content (i.e., temporal fine structure) using the Hilbert transform.

Figure 2.1 shows the schematic decomposition of a speech sound into four frequency subbands (centered at 0.1, 0.4, 1.9, and 8 kHz) using a bank of bandpass filters. Each frequency subband can be further decomposed into its temporal and spectral content (e.g., by using the Hilbert transform; Smith et al., 2002).The temporal structure of the signal is described by the *temporal envelope*, which characterizes the slow amplitude fluctuations in the signal. The spectral content is described by the *temporal fine structure*, which captures the fast fluctuations in the signal (e.g., pitch information; Rosen, 1992).

In general, vocoding discards the temporal fine structure information while the temporal envelope remains largely preserved. To this end, the temporal envelope in each frequency subband is used to modulate an artificially generated carrier signal (Xu et al., 2005). For *noise-vocoding* (Study 4), the carrier is random noise filtered with the bandpass filter of the respective frequency subband. For *tone-vocoding* (Studies 1&2), the carrier is a pure tone at the filter center frequency. Summation over all vocoded subbands results in a signal with degraded spectral information but preserved temporal information (Shannon et al., 1995).

In order to parametrically vary the degree of spectral degradation, two methods were applied in this thesis. In Studies 1&2, the number of frequency subbands was fixed but the signal was degraded in a variable number of these subbands through tone-vocoding (Hopkins et al., 2008). In Study 4, the number of subbands varied parametrically and noise-vocoding was used to degrade spectral information in all subbands (Erb et al., 2012). Vocoding reduces spectral information in a controlled way, which has been found to increase the difficulty of speech understanding (e.g., Faulkner et al., 2001; Obleser et al., 2008; Sheldon et al., 2008), especially in the presence of background noise (e.g., Hopkins and Moore, 2009).

2.3 Magneto- and Electroencephalography

A characteristic feature of human speech signals is their rich temporal structure. Thus, the investigation of the neural dynamics of speech understanding requires neuroimaging methods with a high temporal resolution. Magneto- and electroencephalography (M/EEG) record brain activity noninvasively with a temporal resolution at the order of milliseconds (Vrba and Robinson, 2001; Malmivuo, 2012; Jackson and Bolger, 2014). In this section, I will describe the neurophysiological basis of M/EEG, as well as different methods for the analysis of M/EEG recordings.

2.3.1 Neurophysiological basis of EEG and MEG

The EEG measures electric potentials with electrodes placed on the scalp surface. The first EEG recordings in humans date back to the 1920s (Berger, 1931). The MEG measures magnetic fields with highly sensitive sensors placed close to the scalp. Human MEG was first recorded in 1968 (Cohen, 1968). The main sources of the human M/EEG are postsynaptic potentials of cortical pyramidal cells. The M/EEG signal does not directly reflect spiking activity of nerve cells but rather a smoothened version of the local field potential (Buzsaki et al., 2012)

Since cortical pyramidal cells are organized in a parallel fashion, postsynaptic potentials of huge cells assemblies can sum up, which eventually results in measureable signals in the M/EEG (Hillebrand and Barnes, 2002). The intracellular (i.e., *primary*) currents generate extracellular (i.e., *secondary*) currents which are deflected by the surrounding tissue. While the EEG measures these secondary currents, the MEG directly measures the magnetic fields perpendicular to the primary currents (Lopes da Silva, 2013). The EEG signal is further spread (i.e., smeared) by the limited tissue conductivity of skull and scalp. Contrary, the magnetic signal in the MEG is largely unaffected when it passes the head tissue (Leahy et al., 1998). Compared to the EEG, MEG signals are thus more focal and less spread in space, which allows for a more precise localization of the underlying neural generators (for details, see also Study 3).

The folding of the cortex in sulci and gyri affects the direction of the electric currents generated by the cortical pyramidal cells. The EEG is sensitive to sources that are oriented radially and tangentially with respect to the scalp. The MEG is only sensitive to tangential sources (Ahlfors et al., 2010). Another important difference between EEG and MEG is their sensitivity to deep and shallow sources. In principle, the EEG is also sensitive to deeper sources whereas MEG recordings from planar gradiometer sensors suppress activity from distant sources and are only sensitive to shallow sources in the cortex directly below them (Hämäläinen, 1995).

2.3.2 Analysis of evoked and induced activity

Detailed descriptions of EEG and MEG recording protocols and analysis techniques can be found in the methods sections of Studies 1–3. Here I describe a general division into two different methods used to analyse M/EEG data in the present thesis, namely the analysis of *evoked* and *induced* activity. Evoked and induced activity do not refer to different types of data, but rather to different ways of analysing the same data (for a comprehensive description of evoked and induced activity, see also Tallon-Baudry and Bertrand, 1999).

Figure 2.2 contrasts the analysis of evoked and induced activity for EEG data of a single participant recorded in a listening task. In general, evoked activity is phase- and time-locked across individual trials. The most common way to analyse evoked activity is the event-related potential (ERP; e.g., Kutas and Hillyard, 1980; see also Study 1), which is calculated by averaging across single trial time-domain data. Activity that is not strictly phase- and time-locked across trials is considered noise and gets suppressed in the ERP analysis. The signal to noise ratio (SNR) increases with the square root of the

number of trials. Two more sophisticated methods used for the analysis of evoked activity in the present thesis are inter trial phase coherence (ITPC; Lachaux et al., 1999; see also Study 3) and cross-correlation of EEG signal and the stimulus (e.g., Horton et al., 2013; see also Study 2.2). An important limitation in the analysis of evoked activity is the fact that all activity that is not strictly consistent across trials is cancelled out.

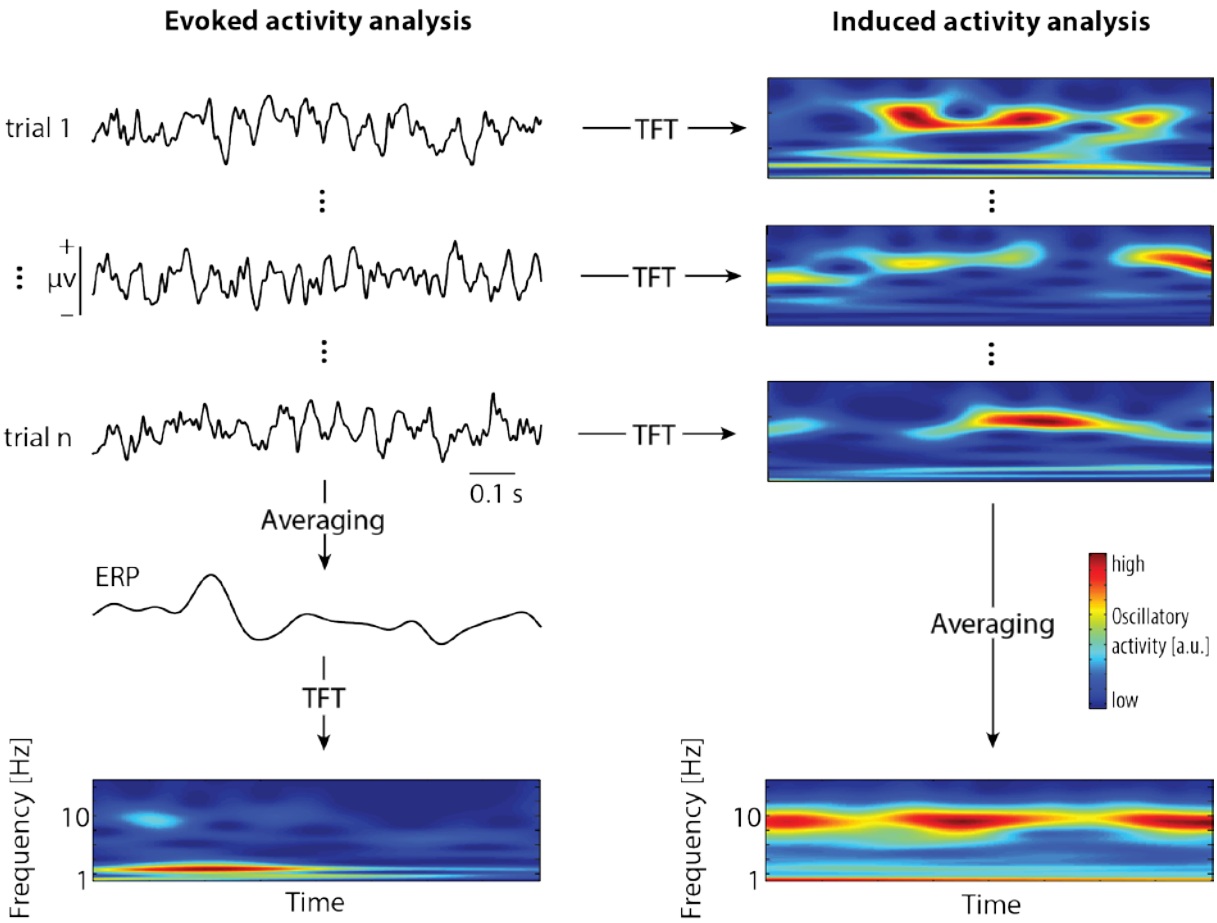

Figure 2.2. Analysis of evoked and induced activity. Left side: analysis of evoked activity of EEG data from a single participant. Single trials are averaged in the time-domain to calculate the event-related potential (ERP). Three representative trials out of 47 trials used for the analysis of the ERP are shown. Time-frequency transformation (TFT) of time-domain data yields oscillatory power (convolution of the ERP with a family of morlet wavelets, using the Fieldtrip toolbox (Oostenveld et al., 2011) for Matlab; wavelet width: 7 cycles; frequencies: 1–15 Hz in steps of 0.1 Hz; time: 1.6 seconds in steps of 0.01 s). Right side: analysis of induced activity for the same dataset. Time frequency transformations of individual trials were performed and averaged across trials. Note the striking difference in the time frequency representations of evoked (bottom left) and induced activity (bottom right). The analysis of evoked activity emphasizes phase- and time-locked activity in lower frequencies (~2–4 Hz), whereas the analysis of induced activity emphasizes non-phase-locked activity that is strongest in the alpha frequency range (~10 Hz) in this dataset. Blue colors indicate low, red colors indicate high oscillatory activity.

The main rationale in the analysis of induced activity is that voltage fluctuations that are not phase- and time-locked across trials are not considered noise but meaningful signal (e.g., Klimesch et al., 1998). To analyse induced activity, the time-frequency transformation (TFT) of time-domain data yields oscillatory power for each trial (e.g., by convolution of the time-domain signal with a family of morlet wavelets; for details, see Studies 2&3). Figure 2.2 shows strong activity in the alpha frequency

range (~10 Hz) in single trials. Since alpha activity is not phase- and time-locked across trials, it gets cancelled out in the analysis of evoked activity (Figure 2.2, bottom left) but is preserved in the analysis of induced activity (Figure 2.2, bottom right).

3 Study 1: Allocation of attention in the face of degraded acoustics

This study investigates event-related potential (ERP) signatures of attention allocation to speech in background noise. Study 1.1 compares neural mechanisms of attention in younger and older listeners in an auditory number comparison task. Study 1.2 further explores in how far increasing task-relevance through monetary incentives affects listeners' allocation of attention.

3.1 Study 1.1: Acoustic detail guides attention allocation in a selective listening task[3]

3.1.1 Introduction

Listening to one talker despite distracting speakers ("cocktail party problem"; Cherry, 1953) requires selective attention, that is, preferential processing of a specific signal at the expense of distractor signals (Kerlin et al., 2010). The demand on selective auditory attention is particularly high if listening conditions are compromised because of hearing loss (Tun et al., 2009) or signal degradation (Wild et al., 2012). It is unknown how and to what extent listeners of different age retain the ability to flexibly allocate attention to changing stimulus acoustics. Here, the electroencephalogram (EEG) was recorded in order to trace neural signatures of selective attention deployment, while younger (20–30 years) and older (60–70 years) healthy listeners performed an effortful selective listening task, in which varying degrees of acoustic degradation implicitly signalled task difficulty.

Fluctuations in cortical excitability have been proposed to regulate auditory selective attention (Schroeder and Lakatos, 2009; Lakatos et al., 2013b), by lowering sensory thresholds for relevant stimuli. Cortical excitability is enhanced by the depolarization of pyramidal neurons, causing slow cortical potentials of negative amplitude in the EEG (He and Raichle, 2009). One well-studied slow potential is the contingent negative variation (CNV; Walter et al., 1964) which occurs after a warning signal during the anticipation of an imperative stimulus (e.g., Zanto et al., 2011b; Chennu et al., 2013). The CNV magnitude is lowered when participants' selective attention to task-relevant stimuli is impaired by distractors (McCallum and Walter, 1968; Tecce and Scheff, 1969; Travis and Tecce, 1998). In turn, larger CNV magnitudes at stimulus onset improve detectability of visual (O'Connell et al.,

[3] This section is adapted from the article published in the Journal of Cognitive Neuroscience by Wöstmann, Schröger, & Obleser (in press).

2009) and auditory targets (Rockstroh et al., 1993). These findings suggest that CNV magnitude correlates with selective attention, possibly through an enhancement of excitability in task-relevant cortical neural networks (Raichle, 2011). It is thus a timely endeavour to exploit the CNV for a refined understanding of selective auditory attention in younger and older listeners.

To study the CNV in a well-controlled, nonetheless ecologically valid selective listening situation, participants performed an auditory number comparison task (Moyer and Landauer, 1967) masked by a distracting talker. To vary the effort of selective attention (Shinn-Cunningham and Best, 2008), perceptual separability of digits and masker was altered by parametrically degrading temporal fine structure (TFS; Moore, 2008b); an acoustic feature found highly relevant for listening against fluctuating maskers (Hopkins and Moore, 2009, 2010). Critically, the onset of the masker served as a warning stimulus in the present design, since the degree of acoustic degradation in the masker implicitly signalled task difficulty and allowed a graded allocation of attention to compensate for unfavourable acoustic conditions. Thus, the dependent neural measure in the present study was the CNV evoked by the onset of the speech masker.

In this attention-demanding selective listening task, we expected improved performance with more preserved acoustic detail. Decreased CNV magnitude with more acoustic detail would indicate that participants adaptively allocate less attention as the signal quality improves. To further tighten the link between the CNV and mechanisms of auditory attention, we anticipated, first, absent or reduced CNV modulation in a control experiment when acoustic detail would not cue task difficulty, and second, a correlation between CNV magnitude and a behavioural marker of individual attentional capacity. Through careful adjustments of stimulus intensities to participants' individual requirements, we were able to investigate the neural mechanisms of auditory attention allocation independent of age-differences in signal audibility or overall performance level. We asked whether healthy aging would affect the flexible allocation of attention to changing acoustic conditions.

3.1.2 Materials and methods

3.1.2.1 Participants

Twenty younger (age range, 20–30 years; mean age, 25.7; 9 females) and twenty older (age range, 60–70 years; mean age, 64; 11 females) healthy, right-handed German native speakers participated in the main experiment. Data of 38 participants were included in the final analysis (see below). Participants

gave informed consent and were financially compensated for participation. Procedures were approved by the local ethics committee of the University of Leipzig Medical faculty.

3.1.2.2 Speech materials

German spoken digits from 21 to 99 (excluding multiples of ten) were recorded from a trained female speaker (sampling rate, 44.1 kHz). All digits contained four syllables and had an average length of 1.125 s (*SD* = 0.056 s). The distracting masker was extracted from a German audiobook (Oscar Wilde, "Der junge König") spoken by a female talker (sampling rate, 44.1 kHz). To increase the energetic overlap of masker and spoken digits, silent periods longer than 70 ms were removed automatically from the masker (using a customised Matlab script R2013a; MathWorks). The resulting audio file had a length of 29'52", from which we extracted 1000 random snippets with a length of six seconds.

For each stimulus, two spoken target digits (referred to as S1 and S2) and one masker snippet (referred to as masker) were selected randomly. Intensities of digits and masker were modified to realise different *Target-to-Masker Ratios* (TMRs; which were individually titrated, see below). For this purpose, root-mean-squared (RMS) masker intensity was fixed at –30 dB full-scale (dBFS) while digit intensity was further reduced (using the *AttenuateSound* function from the psychoacoustics toolbox for Matlab). For example, for a TMR of –15 dBFS, and given the masker intensity of –30 dBFS, intensities of S1 and S2 were set to root-mean-squared –45 dBFS. Lastly, digit and masker signals were combined.

To modify the amount of acoustic detail (temporal fine structure, TFS), the combined signal (composed of masker and digits) was divided in frequency space into 16 overlapping channels (using a gammatone filterbank implemented in the auditory toolbox for Matlab; Slaney, 1993). Channel centre frequencies increased exponentially from 0.08 to 10 kHz. TFS was preserved in all channels below and including six TFS preservation cut-offs (0, 0.11, 0.21, 0.4, 0.76, and 1.45 kHz) and degraded above (Figure 3.1A). Thus, TFS was always degraded in channels above 1.45 kHz but was systematically degraded across conditions in channels at and below 1.45 kHz. We did not preserve TFS above 1.45 kHz, as we observed the largest performance increase up to this frequency in a behavioural pretest (n = 12). All channels below and including the TFS preservation cut-off were left unchanged (i.e., "intact"; Lorenzi et al., 2006). In channels above a given TFS preservation cut-off, the speech envelope was extracted using the Hilbert transform (Smith et al., 2002). The envelope was used to modulate a sinusoidal tone with random starting phase at the channel centre frequency. The resulting signal was

filtered again with the initial filters to remove out-of-channel frequency components (Lunner et al., 2012). The RMS amplitude of the signal in each channel was equalized to this channel's RMS after initial filtering. Finally, intact and modified channels were combined, yielding six different TFS preservation levels. Note that a TFS preservation of 0 kHz meant that TFS was entirely degraded in all 16 channels (Figure 3.1A, top panel) whereas a TFS preservation of 1.45 kHz meant that TFS was preserved in channels below and including 1.45 kHz and was degraded in all channels above (Figure 3.1A, bottom panel).

In essence, our manipulation substantially degraded the fast spectro-temporal fluctuations in higher frequencies, while leaving the slow temporal envelope fluctuations largely intact (Shamma and Lorenzi, 2013). Lower levels of TFS preservation made the signal sound tinny and artificial, rendering perceptual segregation of masker and digits perceptually more demanding. Importantly, speech with degraded TFS in all channels ("vocoded speech") is intelligible if presented in quiet, provided the number of channels is sufficiently high (Shannon et al., 1995; Obleser et al., 2007; Obleser et al., 2008).

3.1.2.3 Hearing acuity

To assess an objective measure of hearing acuity, participants' pure-tone air-conduction audiometric thresholds (at frequencies: 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz) were assessed by a trained audiologist separately for both ears in steps of 5 dB HL using a clinical audiometer (according to the procedures described in: British Society for Audiology, BSA, 2011). Participants did not show interaural asymmetries (≥ 20 dB difference between both ears at more than two frequencies). Hearing thresholds of younger and older participants are shown in Figure 3.1B. Notably, none of the participants were using a hearing aid, nor were any of them subjectively aware of significant hearing impairments.

3.1.2.4 Individual adjustments of materials

One of the main rationales of the present study was to investigate the effect of acoustic signal or age on attention allocation while controlling for potentially confounding between-subject differences in signal audibility or overall task performance level. Prior to the actual experiment, we thus adjusted stimulus intensities to individual requirements to assure a comparable level of task performance across (younger and older) participants on stimulus materials under the most severe degradation (TFS preservation of 0 kHz). In the EEG experiment, we then systematically enhanced the degree of preserved TFS in stimulus materials. We explored in how far younger and older participants'

behavioural responses and neural markers of attention allocation were sensitive to these changes in the degree of TFS preservation.

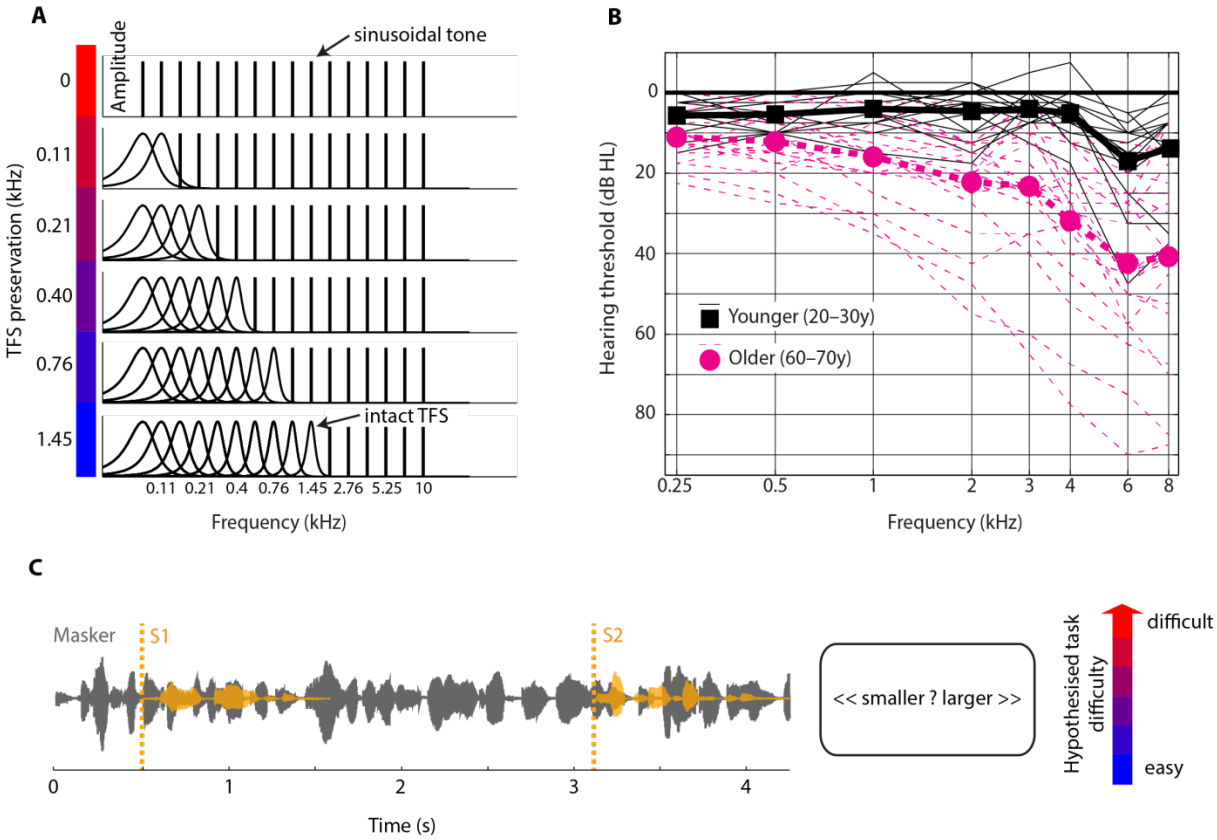

Figure 3.1. Manipulation of acoustic detail and hearing thresholds. (A) Schematic illustration of temporal fine structure (TFS) manipulation. TFS was preserved in channels at lower frequencies and replaced by sinusoidal tones at higher frequencies. The TFS preservation level (highest channel with intact TFS) varied over six levels (0, 0.11, 0.21, 0.4, 0.76, and 1.45 kHz). Amplitudes in all channels are equalised for illustration purpose only. (B) Pure-tone air-conduction audiometric thresholds averaged over both ears for 18 younger (black) and 20 older (magenta) participants. Thick lines show average thresholds for the two age groups. (C) Trial design of the auditory number comparison task. Two spoken digits (S1 and S2) were presented against a distracting speech masker. Task difficulty was hypothesised to increase with lower levels of TFS preservation.

First, to equate audibility of materials despite considerable inter-individual differences in hearing thresholds (Figure 3.1B), overall stimulus intensity was adapted to hearing abilities. To this end, a frequency-specific amplification based on hearing thresholds from 0.25 to 6 kHz was applied to all materials using the CAMEQ procedure (Moore et al., 1998). In essence, this procedure raises signal intensities at frequencies that showed elevated hearing thresholds.

Second, since performance levels in auditory tasks cannot be matched between age groups by controlling for pure-tone audiometric thresholds alone (see Pichora-Fuller et al., 1995), we individually adjusted the Target-to-Masker Ratio (TMR; Schneider et al., 2000). To this end, we varied the TMR systematically while participants performed the auditory number comparison task on materials without preserved TFS ("0 kHz") in an adaptive tracking procedure (two-down one-up

procedure; targeting approx. 71 % accuracy; Levitt, 1971). Testing started at a favourable TMR of +10 dB. This made it rather easy for all participants to perform the number comparison task initially. After two successive correct trials, TMR was decreased (two-down), reducing intelligibility of digits. After one incorrect trial, TMR was increased (one-up). Younger participants performed three and older participants four sessions of adaptive tracking. The individual TMR used in the actual experiment was estimated from the average results of all tracking sessions.

3.1.2.5 Processing speed

Processing speed was assessed with a standard visual test for attentional capacities (d2–R; Brickenkamp et al., 2010). Participants had to mark target letters in twelve lists containing targets and highly similar non-targets. They were instructed to perform the task "as fast and as accurately as possible" and were given 20 s to work on each list, after which they were prompted to switch immediately to the subsequent list. As a test score, we calculated the sum of processed targets on all lists ("BZO" score; possibly ranging between 0 and 308) with high scores indicating high processing speed (Bates and Lemay, 2004).

3.1.2.6 Working memory

Working memory capacity was assessed with the auditory backward digit span test (subtest of the Wechsler Adult Intelligence Scale-Revised; Wechsler, 1984). On each trial, participants were presented a list of spoken digits between one and nine. Digits were spoken by a female voice at a rate of approximately one digit per second and presented at ~75 dB SPL. Participants' task was to repeat the digits in reverse order. The test had seven levels with list lengths increasing from two to eight digits. Each level comprised two items. Participants' responses were marked as correct only if all digits were repeated in the correct order. Testing stopped when participants performed incorrectly on both items for a particular list length. The individual backward digit span score (also referred to as "BSpan") was calculated as the sum of correctly completed items, possibly ranging between 0 and 14.

3.1.2.7 Experimental procedure

Participants were instructed to perform the number comparison "as fast and as accurately as possible". Each trial started with the presentation of the two response options ("kleiner", smaller; "größer", larger) on a computer screen. Auditory stimulation with the manipulated speech masker started after 1.5 s. Spoken digits (S1 and S2) were placed 0.5 s and 3.125 s after masker onset, respectively, resulting in an average delay interval of 1.5 s between S1 offset and S2 onset. All audio files ended

simultaneously with S2 offset and had a length of ~4.25 s (Figure 3.1C). Participants indicated via button press on a response box whether the second digit was smaller (left button pressed with left thumb) or larger (right button pressed with right thumb) than the first. Next, they rated their confidence in this response on a three-point scale (1 = unconfident; 3 = confident). The next trial started self-paced with an additional button press. Behavioural data were recorded by Presentation software (Neurobehavioral Systems).

Each participant performed 300 trials, 50 for each TFS preservation level. For each trial, it was determined randomly whether the second digit was in fact smaller or larger than the first. The experiment was divided in five blocks. Each block contained ten trials for each TFS preservation level in random order, meaning that the level of TFS preservation changed from trial to trial. Blocks were separated by short breaks. The experiment lasted approximately 70 minutes.

3.1.2.8 Behavioural data analysis

Data from two younger participants were excluded from all analyses because of technical problems during data acquisition and below-chance performance (38 % correct), respectively.

To analyse differences in the individual adjustments of materials between age groups, the effect of Age group on individually-titrated TMR was analysed with an independent-samples t-test. The relationship between Working memory and TMR was analysed using a Pearson's correlation (Figure 3.2B).

To quantify participants' performance in the auditory number comparison task, accuracy on each trial (correct vs incorrect) was weighted by confidence ratings to get a more fine grained measure of task performance (Kitayama, 1991). As a result, correct responses were transformed to 100 % weighted accuracy in case of high confidence ratings, to 80 % in case of medium confidence, and to 60 % in case of low confidence. Similarly, incorrect responses yielded 40 % weighted accuracy for low confidence ratings, to 20 % for medium confidence, and to 0 % for high confidence ratings. In the remainder of this paper we use, for simplicity, the term 'accuracy' to refer to accuracy weighted by confidence ratings. As a second measure of task performance, we analysed participants' response times in the number comparison task. In detail, response times corresponded to the time interval between the onset of the second digit and participants' button press to indicate whether the second digit was smaller or larger than the first.

For statistical analyses, we calculated *linear coefficients* characterizing the linear change (slope) of accuracy and response times over the six levels of TFS preservation for each participant (predictor values: –2.5, –1.5, –0.5, 0.5, 1.5, 2.5; using the *polyfit* function in Matlab). In order to test for significant effects of TFS preservation on performance measures, the distribution of linear coefficients was tested against zero (using a one-sample t-test). To test for effects of Age group, we compared younger and older participants' linear coefficients, overall (condition-independent) accuracy measures, and overall response times (using independent-samples t-tests).

3.1.2.9 EEG recording and analyses

Electroencephalography (EEG) was recorded at a 500-Hz sampling rate with a DC–135 Hz pass band (TMS international, Enschede, The Netherlands). Twenty-eight electrodes (Ag/Ag-Cl) were placed at the following positions (Easycap, Herrsching, Germany): Fpz, Fp1, Fp2, Fz, F3, F4, F7, F8, FC3, FC4, FT7, FT8, Cz, C3, C4, T7, T8, CP5, CP6, Pz, P3, P4, P7, P8, O1, O2, left mastoid (A1), and right mastoid (A2). The reference electrode was placed at the tip of the nose and the ground electrode at the sternum. The electrooculogram was recorded from vertical and horizontal bipolar montages. All electrode resistances were kept below 5 kΩ.

Offline, data were analysed using Matlab and the Fieldtrip toolbox (Oostenveld et al., 2011). Epochs were extracted from the continuous signal around masker onset (–2 to 6.5 s). Epochs were low-pass filtered at 100 Hz and baseline corrected by subtracting the mean amplitude in the time interval –0.1 to 0 s. An independent component analysis (ICA) was performed on the epoched data. Components corresponding to eye blinks, saccadic eye movements, muscle activity, electrode drifts, and heartbeat were identified and rejected by inspection of the components' topographies, frequency spectra, and time courses. Remaining artifact-contaminated trials were deleted after visual inspection of EEG waveforms at all electrodes. On average, 7 ± 1 % (*SE*) of trials were rejected from further analyses. Prior to statistical analyses, data were further low-pass filtered at 20 Hz (fourth-order Butterworth filter, zero phase shift).

To calculate the event-related potential (ERP), the time-locked average over all artifact-free trials (irrespective of whether the number comparison was performed correctly or incorrectly) was computed separately for the six TFS preservation levels for each participant. To detect significant effects of TFS preservation on ERP amplitude, a two-level statistical analysis was applied (cf. Obleser et al., 2012; Wilsch et al., 2014). On the first (individual) level, EEG recordings from all trials at 28 scalp

electrodes and between 0 and 4.25 s (relative to masker onset) were submitted to a parametric regression t-test for independent samples (implemented in the *ft_timelockstatistics* function in Fieldtrip). For this regression, we used linearly spaced zero-centered predictor values (–2.5, –1.5, –0.5, 0.5, 1.5, 2.5), to model the monotonic change of ERP amplitude over six levels of TFS preservation. For each participant, we obtained an electrode–time matrix of linear coefficients characterizing the linear change (slope) of ERP amplitude with increasing TFS preservation.

On the second (group) level, individual matrices of linear coefficients were tested for significant differences from zero using a cluster-based permutation dependent samples t-test (Maris and Oostenveld, 2007). First, this test clustered t-values of adjacent points in electrode–time space with a p-value < 0.05, considering a minimum of three neighbouring electrodes as a cluster. Next, the summed *t*-value of each cluster was computed and compared against the distribution of 1000 iteratively and randomly drawn clusters from permuted-labels data. The cluster *p*-value resulted from the proportion of Monte Carlo iterations in which the summed *t*-statistic of the observed cluster was exceeded. As we performed this analysis as a two-sided test (for clusters exhibiting positive and negative effects), clusters with $p < 0.025$ were considered significant. Linear coefficients significantly larger than zero would indicate that ERP amplitude became more positive with higher levels of TFS preservation. The analysis revealed one extensive significant cluster (Figure 3.3).

To test whether the effect of TFS preservation on ERP amplitude in the significant cluster differed between age groups, individual linear coefficients were averaged over electrodes and time points of the significant cluster and submitted to an independent samples t-test with the between-subjects factor Age group (Figure 3.3C). To directly compare the two Age groups in their exhibited ERP amplitude change with higher levels of TFS preservation during the entire trial (not only in the significant cluster), younger and older participants' individual matrices of linear coefficients were submitted to another cluster-based permutation independent samples t-test (between-subject factor: Age group).

To test whether CNV magnitude in individual trials was related to accuracy in the number comparison task, we performed a median split on single trial CNV magnitude in the significant cluster. We calculated the mean accuracy for trials with a small and large CNV magnitude for each participant and level of TFS preservation (Figure 3.4). For statistical analysis, a repeated-measures ANOVA (within-subject factors: TFS preservation & CNV magnitude; between-subjects factor: Age group) was applied to these data.

We analysed whether the magnitude of the CNV would correlate with neuropsychological markers of individual attentional capacity. To this end, we focused on the early CNV (0.1–0.5 s) prior to S1 onset, which was independent of processing task-relevant digits but thought to reflect the preparatory allocation of attention for the ensuing number comparison task. We correlated overall early CNV magnitude (i.e., averaged over all electrodes of the significant cluster and over all conditions) with d2–R scores for processing speed. To control for a possible confound of entering two different groups of participants (younger and older) in one correlation analysis, we also controlled for the effect of Age group in a partial correlation (Figure 3.6). Effects of Age group on overall early CNV magnitude and d2–R scores were analysed with independent-samples t-tests.

3.1.2.10 Control experiment

In a control experiment, we slightly altered the acoustic processing scheme to obtain masker signals identical to the main experiment, but to preserve the temporal fine structure of the spoken target digits. Masker and target digits were submitted to the TFS manipulation (Figure 3.1A) separately, such that acoustic detail (TFS) was only manipulated in the speech masker (over the same six levels as before) but was always preserved up to 1.45 kHz (i.e., maximally intact) in spoken digits.

We hypothesised that task difficulty would be unaffected by these varying masker signals since the task-relevant digits were always maximally intact. Thus, changing acoustic detail in the masker was expected to be no longer an indicative cue on task difficulty in the control experiment. All other experimental and analysis procedures, however, were identical to the main experiment. Importantly, the acoustic stimulation prior to S1 onset was physically identical in the main and in the control experiment. Therefore, we restricted the analysis of ERP data to the time interval of the early CNV prior to S1 (0.1–0.5 s). We re-invited six (three younger, three older) participants 8–12 months after participating in the main experiment. All six had shown a prominent CNV effect in the main experiment (Figure 3.5A).

For statistical analysis, we computed average linear coefficients for the monotonic change in CNV amplitude with higher levels of TFS preservation before S1 onset at electrode Fz in the main and control experiment for each participant. This allowed us to quantify precisely the effect of acoustic detail on CNV in the individual, which allows for compelling within-subject comparisons despite the comparably low number of participants re-invited for the control experiment. Lastly, distributions of

linear coefficients from main and control experiment were tested against zero (using one sample t-tests) and compared between main and control experiment (using a paired t-test).

3.1.3 Results

3.1.3.1 Individual adjustments of speech materials

Figure 3.2A shows younger and older participants' average Target-to-Masker Ratio (TMR) resulting from the individual adjustments of speech materials. As expected, average TMR was significantly lower for younger compared with older participants (t_{36} = 3.60; p = 0.001), showing that younger participants were able to perform the number comparison task under more compromised acoustic conditions.

Figure 3.2B shows individual TMRs as a function of working memory capacity measured with the backward digit span test. The correlation was significant (r = –0.49; p = 0.002; controlling for Age group: p = 0.018), indicating that participants with a smaller working memory capacity required a higher TMR in the auditory number comparison task. When the correlation was computed separately for the two age groups, it reached significance only for older (r = –0.52; p = 0.018) but not for younger participants (r = –0.03; p = 0.903), showing that the relationship between TMR and working memory capacity was mainly driven by the group of older participants. Generally, younger participants performed significantly better in the working memory test compared with older participants (t_{36} = 2.19; p = 0.035).

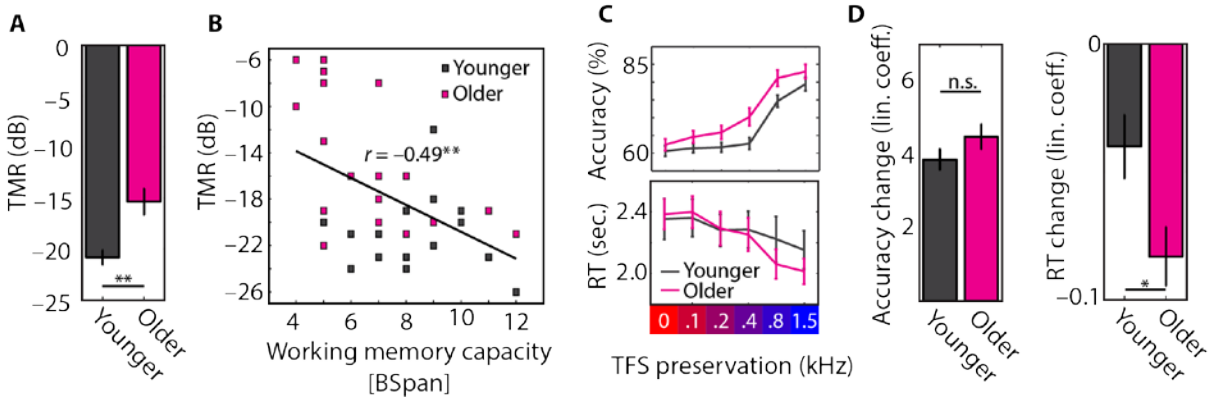

Figure 3.2. Stimulus adjustments and task performance. (A) Average Target-to-Masker Ratio (TMR) used for the individual stimulus adjustments for younger and older participants. (B) Scatterplot of TMR as a function of working memory capacity (auditory backward digit span score; BSpan) for younger (black) and older (magenta) participants. Note that only 33 of 38 data points are visible as some points overlap. (C) Accuracy increased and response times decreased in the auditory number comparison with higher levels of TFS preservation (both p < 0.001). Accuracy was weighted by confidence ratings. (D) Bars show linear coefficients, which quantify the change in accuracy (left panel) and response times (right panel) with each level of TFS preservation. The speed-up of response times with higher levels of TFS preservation was significantly stronger in older participants. *p < 0.05; ** p < 0.01. Error bars show ±1 SE.

3.1.3.2 Performance profits from acoustic detail

Figure 3.2C shows response times and accuracy in the number comparison task for younger (black) and older (magenta) participants. Across age groups, participants showed significantly increasing accuracy (t_{37} = 17.81; p < 0.001) and decreasing response times (t_{37} = –6.95; p < 0.001) as more acoustic detail (TFS) was preserved. The TFS-induced improvement in accuracy did not differ significantly between age groups (Figure 3.2D; t_{36} = 1.35; p = 0.186). Contrary, response times decreased significantly stronger with more TFS in older compared with younger participants (t_{36} = 2.53; p = 0.016). Although Figure 3.2C indicates an overall higher accuracy for older participants, this main effect only approached significance (t_{36} = 1.95; p = 0.059). When we analysed age effects on performance measures separately for un-weighted accuracy values and confidence ratings, we found that better performance in older adults was driven by higher overall un-weighted accuracy (t_{36} = 2.47; p = 0.018) rather than higher confidence ratings (t_{36} = 1.27; p = 0.211). Overall response times did not differ significantly between age groups (t_{36} = 0.28; p = 0.783).

3.1.3.3 Contingent negative variation (CNV) magnitude is modulated by acoustic detail

Figure 3.3A shows the grand average event-related potential (ERP) for six levels of acoustic detail (i.e., TFS preservation). The onset of the speech masker triggered a sustained negative voltage deflection (contingent negative variation, CNV), which was smaller in magnitude for higher levels of TFS preservation. Notably, this CNV magnitude difference was sustained over the entire trial duration and declined after the offset of the acoustic stimulation.

Statistical analysis revealed one significant electrode–time cluster capturing the effect of decreasing CNV magnitude with more acoustic detail in speech materials (p < 0.001; Figure 3.3B). The cluster comprised a large number of mainly fronto-central electrodes and was significant from ~0.1 s up to ~3.8 s after masker onset (Figure 3.3A, grey shaded area). This cluster exhibited a positive effect, indicating that CNV magnitude decreased (i.e., it became more positive in amplitude) with higher levels of TFS preservation. Linear coefficients in Figure 3.3B and C quantify the change in CNV magnitude (in μv) as TFS preservation was enhanced by one level. The effect of TFS preservation on CNV magnitude did not differ significantly between younger and older participants (t_{36} = 0.47; p = 0.639; Figure 3.3C).

One additional positive cluster approached significance (p = 0.036; with α = 0.025 for two-sided testing). This cluster showed a topography similar to the significant cluster (Figure 3.3B) and appeared

in the end of the trial after the significant cluster (3.85–4.25 s). This cluster was not considered in further analyses. No significant clusters exhibiting a significant effect of Age group on ERP amplitude change with higher levels of TFS preservation were found ($p > 0.1$ for all clusters).

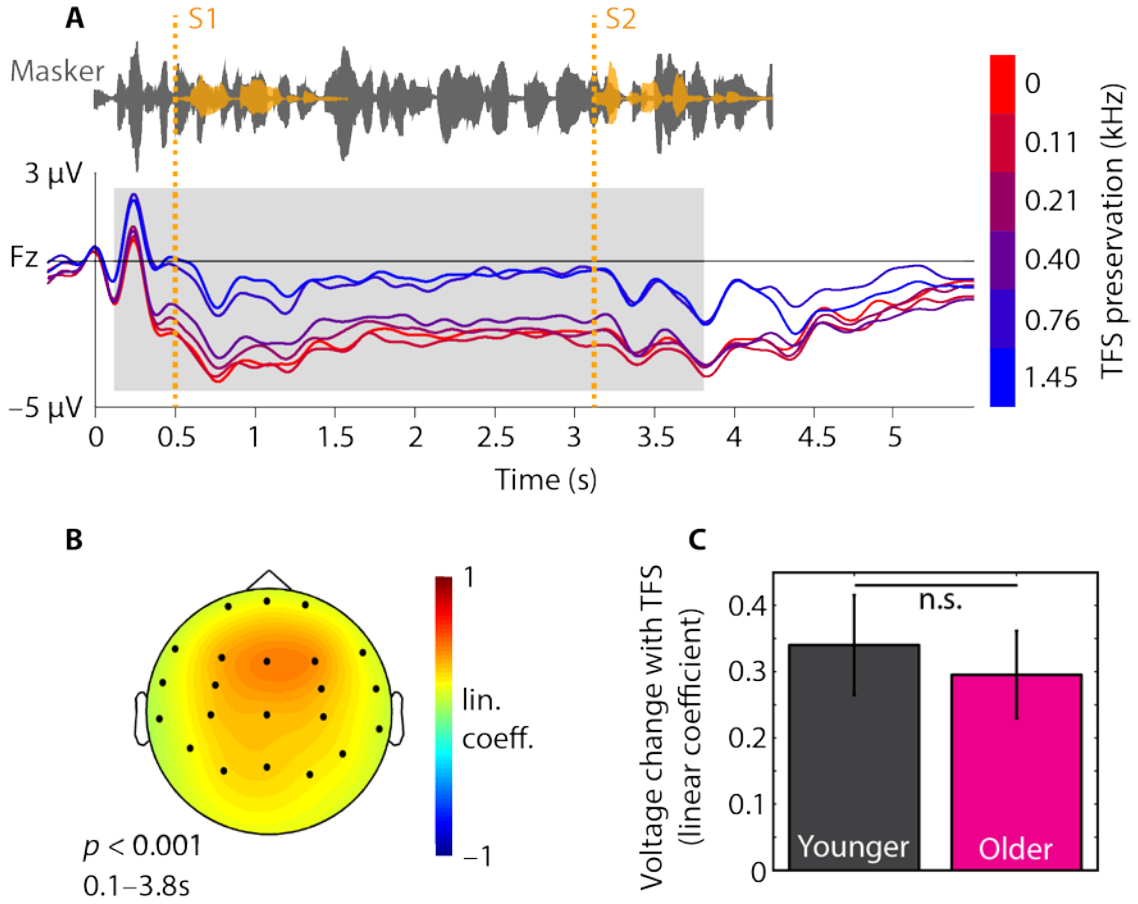

Figure 3.3. Contingent negative variation (CNV) changes with acoustic detail. (A) Schematic trial procedure and grand average event-related potentials (ERPs) at electrode Fz, averaged over all participants for six levels of TFS preservation. ERP waveforms are low-pass filtered at 5 Hz for illustration purpose only. The grey area highlights the time period of the significant increase in ERP amplitude with lower levels of TFS preservation, as revealed by the cluster test (see text). (B) Topography shows average linear coefficients (quantifying the change in ERP amplitude with each level of TFS preservation) for significant electrode–time points in the cluster. Positive linear coefficients indicate that ERP amplitude became larger (i.e., less negative) with higher levels of TFS preservation. Black circles indicate electrodes belonging to the significant cluster (22 of 28 scalp electrodes). (C) Average linear coefficients in the significant cluster did not differ between age groups. n.s., not significant. Error bars show ±1 between-subjects SE.

3.1.3.4 CNV magnitude predicts task performance

Figure 3.4 shows participants' accuracy in the number comparison task separately for trials exhibiting a small or a large CNV magnitude at electrodes and time points of the significant cluster. Across all six levels of TFS preservation, average accuracy was higher in those trials that showed a large CNV compared to trials with a small CNV. Statistical analysis revealed a significant main effect of CNV magnitude on accuracy ($F(1, 36) = 6.67$; $p = 0.014$). This main effect was also significant when we analysed the impact of CNV amplitude on un-weighted accuracy measures ($F(1, 36) = 7.89$; $p = 0.008$)

and confidence ratings ($F(1, 36) = 8.24$; $p = 0.007$) separately. There were no significant two-way or three-way interactions between Age group, TFS preservation, and CNV magnitude (all $p > 0.05$).

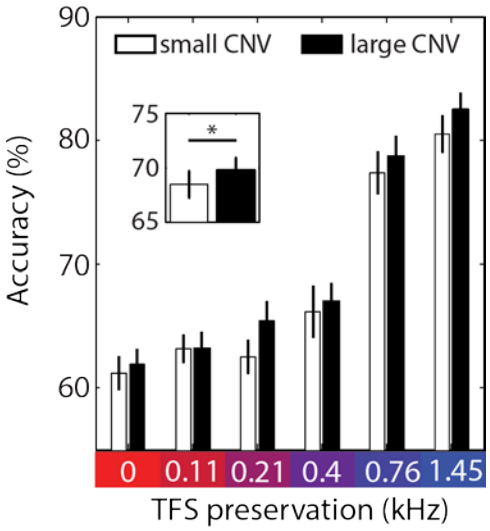

Figure 3.4. Larger CNV amplitude was associated with better task performance. Accuracy in the number comparison task across levels of TFS preservation was higher in trials with large (black bars) compared to small (white bars) CNV magnitude. The inset highlights this main effect of CNV magnitude on accuracy. Accuracy was weighted by confidence ratings. *p < 0.05.

3.1.3.5 Early CNV dynamics and cued task difficulty

An important finding in the present study was that the significant cluster capturing the CNV effect became significant well before the onset of the first digit (S1; Figure 3.3A, grey shaded area). A critical question was whether this early CNV (0.1–0.5 s) was a marker of cued task difficulty or just of the acoustic detail in speech materials. In a control experiment, we thus tested to what degree the early CNV was modulated when acoustic detail was manipulated but cued task difficulty was held constant. To this end, acoustic detail varied only in the masker but was fixed in the target digits. Thus, varying acoustic detail in the masker should not cue task difficulty as task-relevant digits were always maximally intact. For the six participants tested in the control experiment, accuracy did not change with the degree of TFS preservation in the masker ($t_5 = -0.34$; $p = 0.75$; average accuracy = 54 %; average un-weighted accuracy: 61 %) indicating constant task difficulty across conditions.

For the analysis of the early CNV it was critical that the acoustic stimulation prior to S1 was identical in main and control experiment. Thus, any difference in early CNV modulation between main and control experiment could not be due to differences in the acoustic stimulation. Figure 3.5A & B show average CNVs ($n = 6$) for the main and the control experiment, respectively. In the main experiment, early CNV (0.1–0.5 s) magnitude at electrode Fz decreased (i.e., amplitude became more positive) when more TFS was preserved in the speech materials ($t_5 = 12.49$; $p < 0.001$). Crucially, even

in the control experiment, where task demands were constant over conditions, early CNV amplitude decreased with more preserved TFS (t_5 = 4.85; p = 0.005). This finding suggests that the early CNV is sensitive to varying degrees of preserved TFS in the masker even if varying acoustics do not cue task difficulty. Most important for the present study however, the early CNV modulation in the main experiment, where preserved TFS cued task difficulty, was significantly stronger compared to the control experiment (t_5 = 2.92; p = 0.033; Figure 3.5C). In sum, the early CNV is sensitive to acoustic manipulations as such, but it is even stronger modulated if these acoustic manipulations implicitly cue task difficulty.

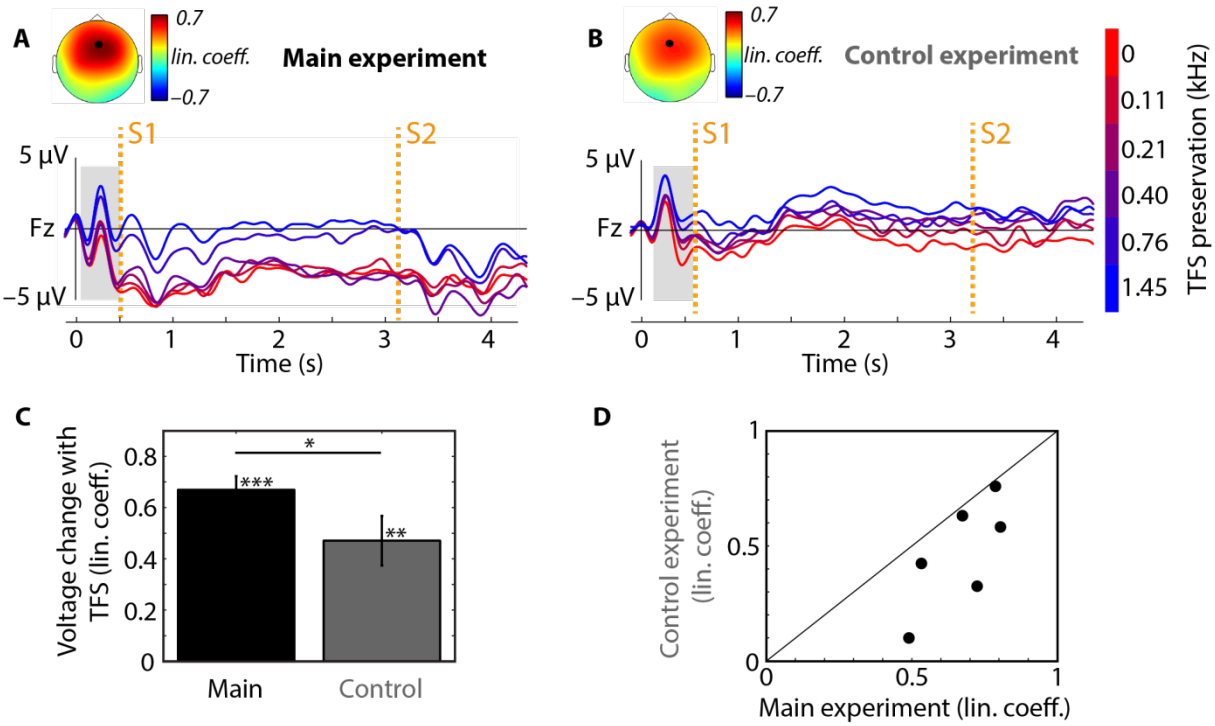

Figure 3.5. Early CNV in main and control experiment (n = 6). (A) Average ERPs of six participants relative to masker onset (0 s) for six levels of TFS preservation in the main experiment, where the acoustic detail in the masker cued task difficulty. (B) ERPs in the control experiment, where acoustic detail in the masker was uninformative about task difficulty. ERP waveforms are low-pass filtered at 5 Hz for illustration purpose only. Grey shaded areas indicate the time interval in which the acoustic stimulation in main and control experiment was identical. This time interval of the early CNV (0.1–0.5 s) was used for statistical analyses. (C) Average linear coefficients capturing the change of early CNV magnitude with higher levels of TFS preservation in the main (black) and control (grey) experiment. Error bars show ±1 SE. (D) Scatterplot of linear coefficients from individual participants in the main and control experiment. Points below the diagonal show that the effect of TFS preservation on the early CNV was stronger in the main compared to the control experiment. **p < 0.01; ***p < 0.001.

Figure 3.5D shows mean linear coefficients, quantifying the change in early CNV amplitude at electrode Fz with higher levels of TFS preservation, for each of the six participants in the main experiment contrasted with the control experiment. The fact that all points fall below the diagonal demonstrates that all six participants showed a stronger CNV modulation in the main compared to the control experiment, indicating the high consistency of this effect across participants.

3.1.3.6 Early CNV magnitude and individual attentional capacity

Lastly, we reasoned that the magnitude of the early CNV reflecting participants' attentional preparation for the ensuing number comparison task should be directly related to individual attentional capacity. Figure 3.6 shows overall (condition-independent) early CNV magnitude (0.1–0.5 s) in the main experiment as a function of d2–R scores for processing speed, an established neuropsychological marker for attentional capacity. The correlation was significant (r = 0.49; p = 0.002; controlling for Age group: p = 0.002), indicating that participants with higher processing speed showed smaller (i.e., more positive) early CNV magnitudes. As is discernible from the scatterplot in Figure 3.6, younger and older participants overlapped largely in both, measures of early CNV magnitude and d2–R scores. Statistical analyses revealed no significant difference of early CNV magnitude between age groups (t_{36} = 0.58; p = 0.568) but a tendency for higher d2–R scores in younger participants (t_{36} = 1.92; p = 0.063).

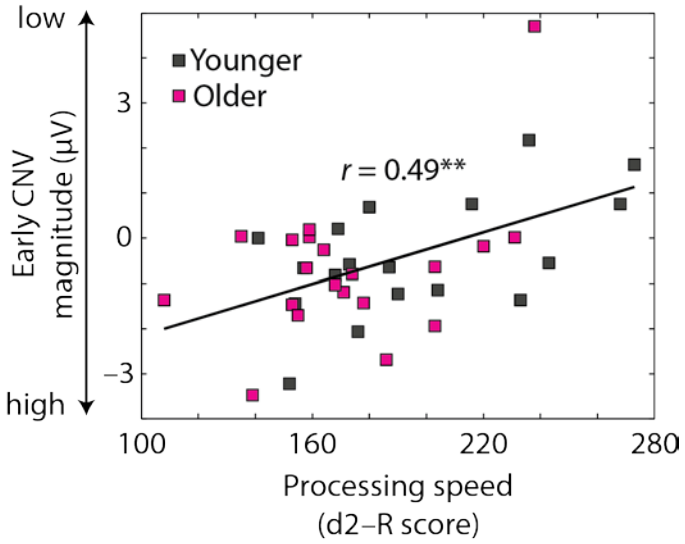

Figure 3.6. Processing speed predicts early CNV magnitude. Scatterplot of overall (i.e., condition-independent) early CNV magnitude (0.1–0.5 s; averaged over all electrodes of the significant cluster) as a function of individual d2–R scores for processing speed for younger (black) and older (magenta) participants. As the CNV is a negative scalp potential, smaller µv values on the y-axis indicate higher CNV magnitude. **p < 0.01.

3.1.4 Discussion

How flexibly can changing acoustics trigger the allocation of attention in a selective listening situation, and how is this attention allocation process affected by healthy aging? Here, we tested the hypothesis that variations in the instantaneous acoustic conditions would signal task difficulty and implicitly cue the allocation of attention in younger (20–30 years) and older (60–70 years) participants. EEG recordings of the contingent negative variation (CNV) served as an index of auditory selective attention.

3.1.4.1 Acoustic detail guides the allocation of attention

The most important finding was a strong dependence of CNV magnitude on preserved acoustic detail (temporal fine structure, TFS) in speech materials. This is a new observation extending previous knowledge on the CNV as a marker of attention allocation: It demonstrates, first, that CNV magnitude is directly and parametrically dependent on the temporal fine structure of the acoustic signal; second, however, this dependency is modulated by the task-relevance of this acoustic cue itself (see control experiment and in-depth discussion below).

As acoustic detail was parametrically preserved from the low frequencies, participants' task performance improved (Figure 3.2C) and CNV magnitude decreased (Figure 3.3). These findings suggest that when the perceptual segregation of digits and masker became less effortful due to more preserved TFS (Hopkins et al., 2008; Moore, 2008b; Hopkins and Moore, 2009, 2010), the task was less attention demanding as reflected in smaller CNV magnitude (McCallum and Walter, 1968; Tecce and Scheff, 1969; Wilkinson and Ashby, 1974; Tecce et al., 1976; Travis and Tecce, 1998; Zanto et al., 2011b; Chennu et al., 2013). On a neuronal level, enhanced CNV magnitude in conditions with less acoustic detail could reflect a lowering of perceptual thresholds through an enhanced cortical excitability in task-relevant cortical networks (Rockstroh et al., 1993; He and Raichle, 2009; O'Connell et al., 2009; Raichle, 2011). In line with this interpretation, combined EEG–fMRI (functional magnetic resonance imaging) studies revealed a positive relationship between blood oxygenation level dependent (BOLD) activity and CNV magnitude (Nagai et al., 2004; Hinterberger et al., 2005; Scheibe et al., 2010), suggesting an enhanced information flow between thalamus and cortex during the CNV period. Our finding of improved task performance in trials with a large CNV magnitude (Figure 3.4) further supports the view that a larger CNV indicates increased selective attention, which, in turn, leads to improved processing of auditory targets embedded in a speech masker.

Figure 3.3A shows that the significant modulation of the event-related potential (ERP) started as early as 0.1 s after masker onset, covering the time range of early auditory evoked potentials (N1, P2; Picton and Hillyard, 1974). Statistical analysis revealed only a single electrode–time cluster exhibiting a significant effect of acoustic detail covering almost the entire trial (foreperiod, target encoding, and retention), as it is typical for slow cortical potentials like the CNV. This finding suggested that the CNV was superimposed on early ERP components and hence, we did not analyse these early evoked potentials in isolation. Instead, we focused largely on the early CNV, emerging right after the onset of the speech masker but before the onset of the first digit (S1). Critically, the early CNV was independent

of processing task-relevant digits, but thought to solely reflect participants' preparation for the number comparison task. In trials with minimal preserved acoustic detail, the speech masker before S1 onset implicitly cued a high task difficulty. Listeners could take advantage of this implicit cue and allocate more selective attention to overcome the unfavourable acoustic conditions. We presumed that the early CNV modulation (0.1–0.5 s) reflected participants' graded allocation of auditory attention as the speech masker implicitly signalled task difficulty.

However, this interpretation implies that the early CNV modulation as a function of acoustic detail should be significantly reduced if acoustic detail in the masker does not cue task difficulty. To test this hypothesis, we conducted a control experiment (Figure 3.5) in which acoustic detail of the masker did not cue task difficulty. In the control experiment, performance did not improve with more acoustic detail showing that task difficulty was unaffected by acoustic detail. Most importantly, the early CNV effect was significantly stronger when acoustic detail cued task difficulty (main experiment), compared to a setting where acoustic detail was uninformative about task difficulty (control experiment). The fact that this pattern of results was consistent over all participants tested in the control experiment (Figure 3.5D), justifies the relatively small sample of six participants in the control experiment. In general, this finding corroborates our conjecture that the early CNV is an indicator of preparatory selective attention allocation triggered by expected task difficulty.

In the control experiment where acoustic detail varied but did not cue task difficulty, the early CNV effect was decreased but not entirely absent. It is thus conceivable, in line with previous research, that degraded acoustic conditions automatically increase the allocation of attention (Obleser and Weisz, 2012; Obleser et al., 2012) even if the degradation applies only to task-irrelevant materials (Winkler et al., 2003). Note that in everyday listening situations, acoustic degradations resulting from reverberations, background noise, or phone lines apply to all transmitted signals (target and masking signals). Therefore, an automatic increase in the allocation of auditory selective attention in adverse acoustic conditions is an effective mechanism to compensate for compromised acoustic conditions.

One important point in our study is to consider whether the observed negative voltage deflection (Figure 3.3) can indeed be considered a CNV. In most classical CNV paradigms, a warning stimulus triggers a negative-going CNV that peaks at the expected time point of a later occurring target stimulus. In our study however, the warning stimulus (masker onset) was followed by two consecutive target stimuli (S1 & S2). The early occurrence of the first target stimulus 0.5 s after masker onset is a possible reason why our negative voltage deflection did not considerably increase in magnitude after

S1 onset. Besides, although the CNV in its narrow sense varies with changing "attention to" or "anticipation of" a target stimulus, our negative voltage deflection was also sensitive to changes of acoustic detail alone (control experiment, Figure 3.5). As described above, we consider it likely that more adverse listening conditions automatically enhanced the allocation of attention, reflected in a stronger negative voltage deflection. Finally, our negative voltage deflection shows a number of properties of typical CNVs since it (1) shows up as a sustained negative voltage deflection strongest over fronto-central electrode sites, (2) is associated with improved task performance if its higher in magnitude (Figure 3.4), and (3) could be directly linked to markers of selective attention (Figure 3.6). Thus, despite the fact that our negative voltage deflection differs slightly from the classical CNV in the narrow sense, we still consider it appropriate to be referred to as a CNV.

3.1.4.2 Early CNV magnitude reflects individual attentional capacities

Evidence for a close relation between individual cognitive capacities and the magnitude of slow cortical potentials (see also Vogel et al., 2005) was given by the significant correlation of overall (condition-independent) early CNV magnitude and the d2–R score for processing speed (Figure 3.6; Brickenkamp et al., 2010). In the d2–R test, visual target items compete with highly similar distractors for limited processing resources (Desimone and Duncan, 1995; Bates and Lemay, 2004). Better participants succeed at selectively attending to targets while ignoring distractors. They can thus process more target items and achieve higher d2–R scores. Here, participants with good selective attention abilities showed smaller (i.e., more positive) overall early CNV magnitudes. Generally, this finding adds weight to the interpretation of the early CNV as a direct electrophysiological index of preparatory selective attention allocation. In particular, this result suggests that the effort of selective attention in a demanding listening task was lower for participants with higher selective attention abilities. In conclusion, the strong link between attentional capacities and CNV magnitude emphasises the importance of taking into account individual cognitive capabilities for the investigation and treatment of subject-specific listening abilities in acoustically demanding situations.

3.1.4.3 Age affects required acoustic conditions and response times

In contrast to prior studies which found age differences both in CNV dynamics (Loveless and Sanford, 1974; Zanto et al., 2011b) and in the accuracy of detecting changes in temporal fine structure (Grose and Mamo, 2010; Hopkins and Moore, 2011), we found age effects rather in the individual adjustments of speech materials required prior to experimental testing and in response times. First, for

several older participants, hearing acuity was reduced (especially at higher frequencies) compared to younger participants (Figure 3.1B). As overall stimulus intensities were adjusted to individual hearing thresholds (CAMEQ procedure; Moore et al., 1998), these older participants were listening to overall more amplified materials during the experiment. Second, older participants required on average a significantly higher Target-to-Masker Ratio (TMR) to reach a similar performance level as younger participants (Figure 3.2A). This result confirms prior research showing that older listeners usually require higher Signal-to-Noise Ratios (SNRs) to hear individual words in noise than do younger listeners (Pichora-Fuller et al., 1995; Murphy et al., 1999; Schneider et al., 2000; Pichora-Fuller, 2003a). The need for less attention-demanding listening conditions in older participants might speak for a decline in attentional control, causing difficulties in attending relevant and ignoring irrelevant sound sources (Passow et al., 2014). Third, the speed-up of response times with higher levels of TFS preservation was stronger in older compared with younger participants (Figure 3.2C&D). Thus, older participants show an enhanced sensitivity to changes in spectral detail (see also Schvartz et al., 2008), implying that older listeners' task performance is particularly dependent on stimulus-inherent features in the acoustic materials. However, as we did not find concomitant differences in CNV dynamics between age groups, it is an open issue for future studies to relate this difference in behaviour to neural changes in the elderly.

The finding that older participants performed poorer in the auditory working memory test (backward digit span) compared to younger participants confirms the general trajectory of decline in memory functioning with age (Salthouse and Kersten, 1993; Fisk and Warr, 1996). More important, however, individual working memory capacity significantly predicted the relative intensity of spoken digits (TMR) determined in the individual adjustments of stimulus materials (Figure 3.2B). Participants with a smaller working memory capacity required more favourable acoustic conditions (higher TMR) to perform the number comparison task. Research has shown that limited resources of the working memory system must be allocated to processing and temporary maintenance and manipulation of speech information (McCoy et al., 2005; Lunner et al., 2009). We presume that participants with fewer memory resources required more favourable encoding conditions to free resources needed for the retention and numerical comparison of digits. In general, this finding demonstrates the tight link between sensory and higher cognitive abilities (Li and Lindenberger, 2002). In sum, aging in and by itself is not critically affecting the ability to allocate attention in a task-adaptive

manner, as long as listening conditions are adjusted to individual sensory acuity and working memory capacity.

3.1.4.4 Conclusions

Dynamics of the early contingent negative variation (CNV) reveal that the instantaneous acoustic conditions in a selective listening task cue the adaptive allocation of auditory selective attention (Fritz et al., 2007) in younger and older listeners. This preparatory allocation of attention for an ensuing task is shown to be partly automatic (driven by characteristics of the signal), but it depends to large extents on the expected task difficulty conveyed by the signal itself (Figure 3.5). The effort of selective attention allocation during the task depended on listeners' individual selective attention abilities (Figure 3.6). Listeners' age is not critically affecting these processes, as long as listening conditions are adjusted to individual sensory acuity and working memory capacity, suggesting that basic mechanisms of preparatory attention allocation are preserved in healthy aging.

3.2 Study 1.2: Influence of monetary incentives on attention allocation

3.2.1 Introduction

In the previous chapter, we found that the magnitude of the contingent negative variation (CNV) was dependent on stimulus-dependent ("bottom-up") acoustic properties of speech materials, but also on participants' task-dependent ("top-down") allocation of attention to overcome expected listening challenges. We concluded that under more degraded acoustics, participants put more effort in the allocation of attention in order to perceive spoken digits in background noise. In this follow-up experiment, we tested more directly whether participants' allocation of attention is reflected in CNV magnitude. We thus varied monetary incentives across trials in the auditory number comparison task under the assumption that higher incentives would enhance participants' deliberate allocation of attention, that is, attentional effort.

In everyday listening situations, the motivation to understand a conversational partner varies as a function of the relevance of the transmitted message. For instance, if a gate change is announced over the loudspeakers at the airport, passengers of the respective flight will follow the announcement more attentively than other people at the airport. Critically, attention is deliberately allocated to the announcement as soon as passengers become aware of the high relevance of the message. Research has shown that attention improves perception of relevant stimuli (e.g., Okamoto et al., 2007; Cohen and Maunsell, 2009; Rotermund et al., 2009). Moreover, it has been shown that enhanced task-relevance through monetary incentive improves performance in attention-demanding tasks (Small et al., 2005; Engelmann and Pessoa, 2007; Zedelius et al., 2012), suggesting that monetary incentives impact attention. However, monetary incentives might also decrease participants' intrinsic motivation and personal interest in a task (for a meta-analysis, see Edward et al., 1999). Thus, we tested whether behavioral performance and a well-known electrophysiological signature of selective attention – the CNV – would be affected by monetary incentives in an effortful listening task.

Monetary incentives are an effective experimental tool to manipulate reward in a performance-related manner (for a review on reward and attentional effort, see Sarter et al., 2006). Neuroimaging studies have revealed that both sub-cortical and cortical brain regions form a functional network that is crucial for the detection of past rewards and the prediction of future rewards (for review, see Schultz, 2000; Knutson et al., 2005). Previous EEG studies revealed that reward magnitude and reward valence affect the amplitude of the P3 and the feedback negativity event-related potential (ERP) components,

respectively (e.g., Yeung and Sanfey, 2004; Sato et al., 2005). Contrary, the effect of monetary incentives on slow cortical potentials such as the CNV is less clear: While some studies found no effect of monetary incentives on CNV magnitude (e.g., Goldstein et al., 2006; Broyd et al., 2012), CNV magnitude was found to increase when participants' motivation was enhanced through higher task difficulty or higher effort to perform a behavioural response (Rebert et al., 1967). This suggests that participants' attentional effort to solve a task at hand might be reflected in the CNV. With this follow-up experiment, we aimed at fostering an understanding of the CNV as a neural signature of attentional effort under varying levels of monetary incentives.

At present, evidence for an effect of monetary incentives on speech comprehension under challenging acoustic conditions is sparse. One recent behavioural study found that increasing listeners' motivation by asking questions concerning the presented speech materials enhanced participants' self-reported listening effort (Picou and Ricketts, 2014). This suggests that a listener's mental state can influence the effort he or she invests to accomplish successful speech comprehension. In this follow-up experiment, we investigated whether monetary incentives would affect behavioural and electrophysiological measures of selective attention to speech under varying acoustic conditions. We expected that higher monetary incentives would increase participants' attentional effort, which would surface in improved behavioural performance and enhanced CNV magnitude.

3.2.2 Methods

The present chapter describes a follow-up experiment of the main experiment discussed in the previous chapter (3.1). Thus, we mention here only those methodological issues that differed from the main experiment.

3.2.2.1 Participants

Nineteen younger participants (age range = 20–30 years; mean age = 24.5; 11 females) took part in this experiment. Participants in this follow-up experiment did not take part in the main experiment (chapter 3.1). The data of five additional younger participants were not used in the data analyses due to a large proportion of artifact-contaminated trials in the EEG recordings (> 50 %).

3.2.2.2 EEG recording and analysis

EEG recordings were carried out using the same EEG system used in the main experiment (chapter 3.1). The only difference from the main experiment was that remaining artifact-contaminated trials

after the rejection of bad components from the independent component analysis (ICA) were not rejected by visual inspection of the data. Instead, all trials in which the range of 140 μv was exceeded in any channel were rejected automatically.

For statistical analyses of the event-related potential (ERP), we applied analyses of variance (ANOVA) instead of cluster-based permutation tests used in the main experiment (chapter 3.1). Cluster-based permutation tests are particularly useful when hypothesis are not specific in time and electrode space. Thus, we used a cluster-based permutation test in the main experiment as we did not know precisely prior to the experiment in which time interval and at which electrodes the CNV effect would occur. For the follow-up experiment however, we had concrete hypotheses concerning the time interval and electrodes contributing to the CNV effect on the basis of the main experiment. To directly test these hypotheses, we applied repeated-measures ANOVAs on CNV amplitude in this follow-up experiment.

3.2.2.3 Procedure

The experimental procedure was the same compared to the main experiment (chapter 3.1) with the following exceptions:

(1) Since we only tested healthy younger participants in this follow-up experiment, we did not expect significant differences in hearing acuity among participants. Thus, we did not adapt the overall stimulus intensity to participants' hearing acuity. However, in line with the main experiment, we adjusted the target-to-masker (TMR) sound level ratio to realize a performance level of ~71 % on materials without preserved fine structure (0 kHz TFS).

(2) Contrary to the main experiment where the preservation of temporal fine structure of speech materials was varied over six levels between 0 and 1.45 kHz (see chapter 3.1), it varied only over three levels (0, 0.4, 1.45 kHz) in the follow-up experiment. Orthogonal to the manipulation of fine structure, we varied the monetary incentive on each trial across three levels (1, 3, 5 euro cent). Each participant completed 270 trials, 30 trials for each one of nine condition in the 3 (fine structure) x 3 (monetary incentive) design. The order of trials was fully randomized. The entire experiment lasted approximately one hour.

(3) Each trial started with a visual cue, indicating the monetary incentive (1, 3, or 5 cent) on this trial. This cue was presented for 1 s prior to the display of the two response options ("kleiner", smaller;

"größer", larger). 1.5 s after the onset of response options, two spoken digits in background speech were presented (Figure 3.7). In the end of each trial, participants received feedback indicating whether they performed the number comparison correctly. In case of a correct response, the incentive was added to participants' balance; in case of an incorrect response, the incentive was subtracted from participants' balance. After the feedback, participants' balance (sum of acquired and lost incentives over the experiment) was presented on the computer screen. After the experiment, participants were financially compensated for their participation with 7 € per hour plus their individual balance acquired over the entire experiment. For participants tested in this study, the balance ranged between 300 and 500 cents.

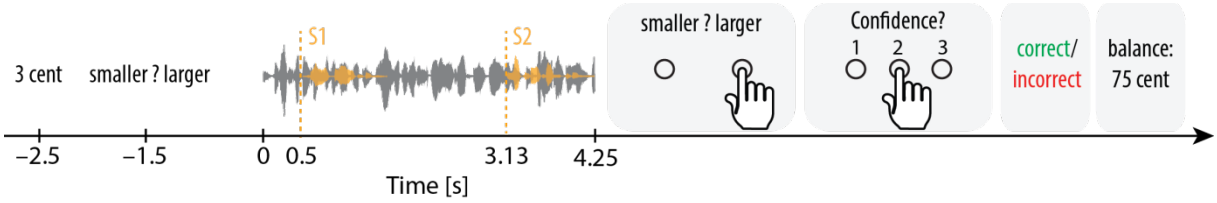

Figure 3.7. Trial design in the follow-up experiment. Each trial started with the presentation of the monetary incentive on the computer screen (1, 3, or 5 cent). Subsequently, the two response options appeared on the screen (smaller, larger), followed after 1.5 s by the acoustic stimulation. Two spoken digits (S1 and S2) were presented against a speech masker. After the acoustic stimulation, participants had to make a decision whether S2 was smaller or larger than S1 and indicate how confident they were in this decision. Participants received feedback about the correctness of their decision and saw their balance (sum of acquired and lost incentives over the experiment) on the screen. For more details concerning the trial design, see also main experiment in chapter 3.1.

3.2.2.4 Effect sizes

To estimate effect sizes for F-statistics (ANOVAs), we calculated the partial eta-squared (η^2_P). Partial eta-squared values of 0.01, 0.06 and 0.14 indicated small, medium, and large effects, respectively (Cohen, 1969). For t-statistics (dependent and independent samples t-tests), we estimated the effect size measure r, which is bound between 0 and 1 (Rosenthal, 1994).

3.2.3 Results

3.2.3.1 Effect of fine structure and monetary incentive on performance

Figure 3.8 shows accuracy and response times in the number comparison task as a function of fine structure in speech materials (in main- and follow-up experiment) and monetary incentives (in the follow-up experiment). First, we tested whether fine structure and monetary incentives affected performance in the follow-up experiment. We conducted two repeated-measures ANOVAs with the factors fine structure (0, 0.4, 1.45 kHz) and monetary incentive (1, 3, 5 cent) on accuracy and response times. In line with the main experiment (chapter 3.1), the main effect fine structure was significant for accuracy ($F(2, 36) = 75$; $p < 0.001$; $\eta^2_P = 0.81$) and response times ($F(2, 36) = 4.45$; $p = 0.019$; ; $\eta^2_P =$

0.2), indicating higher accuracy and faster responses with more preserved fine structure, respectively. Neither the main effect monetary incentive nor the fine structure × monetary incentive interaction was significant for accuracy or response times (all $p > 0.5$; all $\eta^2_P < 0.04$).

Second, we tested whether the effect of fine structure on performance differed between main and follow-up experiment. To this end, we averaged across all levels of monetary incentives in the follow-up experiment and submitted behavioural measures from both experiments to two repeated-measures ANOVAs for accuracy and response times (within-subject factor: fine structure; between-subject factor: experiment). The main effect experiment was not significant for accuracy or response times (both $p > 0.7$; both $\eta^2_P < 0.01$), indicating similar overall performance in main and follow-up experiment. The fine structure × experiment interaction for accuracy approached significance ($F(2, 70) = 2.95$; $p = 0.059$; ; $\eta^2_P = 0.08$). However, post-hoc independent-samples t-tests revealed no significant difference between accuracy in main and follow-up experiment for any level of fine structure (all $p > 0.3$; all $r < 0.18$). The fine structure × experiment interaction for response times was not significant ($F(2, 70) = 0.36$; $p = 0.7$; $\eta^2_P = 0.01$).

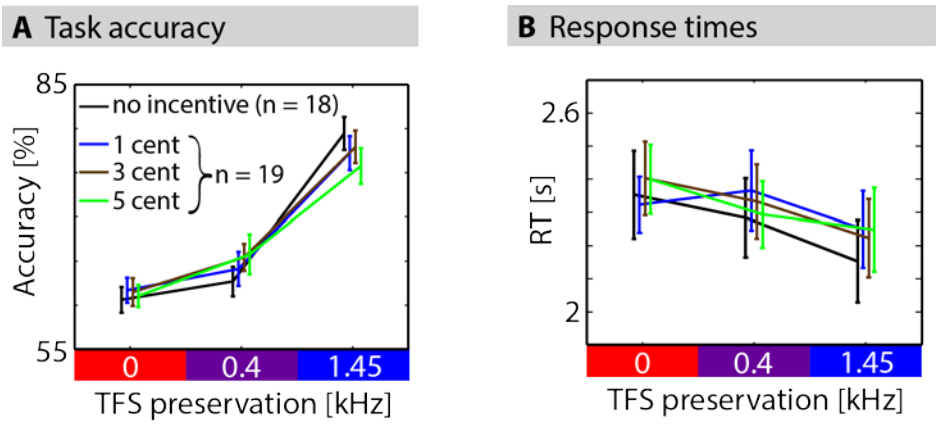

Figure 3.8. Behavioural performance in main and follow-up experiment. (A) Accuracy in the auditory number comparison task as a function of temporal fine structure (TFS) preservation and monetary incentives. The "no incentive" condition shows performance for the group of younger participants (n = 18) in the main experiment, where no manipulation of monetary incentives was applied. The 1, 3, and 5 cent conditions show performance for a different sample of younger participants (n = 19) tested in the follow-up experiment, where monetary incentives were varied. Accuracy data were weighted by confidence ratings. (B) Response times (relative to the onset of the second digit) in the main experiment (no incentive, black) and follow up experiment (1 cent: blue, 3 cent: brown, 5 cent: green). Error bars show ± 1 SEM.

3.2.3.2 Modulation of the contingent negative variation (CNV)

Figure 3.9A&B show grand average waveforms of the event-related potential (ERP) at electrode Fz in the main and follow-up experiment, respectively. A dominant negative voltage deflection (contingent negative variation, CNV) was present in both experiments, starting early after the onset of the distracting speech masker (0.1 s) and lasting until the offset of acoustic stimulation (4.25 s) in the main

experiment, and even longer in the follow-up experiment. Topographic maps in Figure 3.9C show that the modulation of CNV amplitude with different levels of fine structure was most prominent over fronto-central electrodes. For statistical analyses, we calculated the average CNV amplitude (across time points from 0.1 to 4.25 s at electrode Fz) for all participants in main and follow-up experiment.

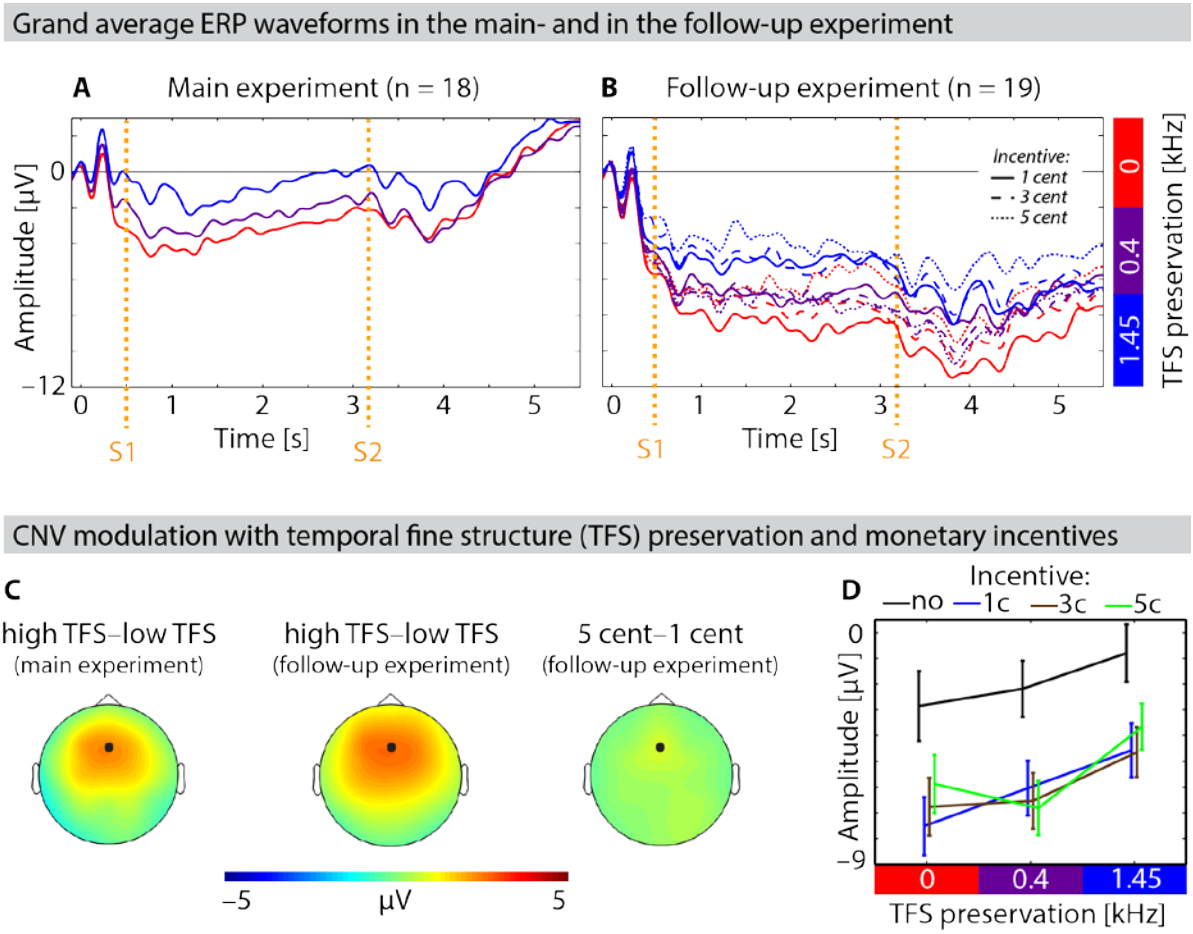

Figure 3.9. CNV in main and follow-up experiment. The grand average event-related potential (ERP) at electrode Fz showed a pronounced negative voltage deflection (contingent negative variation, CNV) for younger participants in the main experiment (A, n = 18) and for a different sample of younger participants in the follow-up experiment (B, n = 19). (C) Topographic maps show average CNV amplitude between 0.1 s after masker onset until the offset of the acoustic stimulation (4.25 s) for the effect of temporal fine structure preservation as the contrast: high TFS (1.45 kHz) – low TFS (0 kHz) and for the effect of monetary incentives as the contrast: 5 cent – 1 cent. Note that the effect of monetary incentive is only presented for the follow-up experiment, as monetary incentives were not manipulated in the main experiment. Electrode Fz used for statistical analyses (see text) is highlighted in black. (D) Average CNV amplitude (averaged across 0.1–4.25 s at electrode Fz) as a function of temporal fine structure (TFS) preservation and monetary incentive. Error bars show ± 1 SEM.

First, we tested the effects of fine structure and monetary incentive on CNV amplitude in the follow-up experiment. To this end, we submitted average CNV amplitudes to a repeated-measures ANOVA with the factors fine structure and monetary incentive. The main effect fine structure was significant ($F(2, 36) = 14{,}67$; $p < 0.001$; $\eta^2_P = 0.45$), indicating larger CNV magnitude (i.e., stronger negativity) when less fine structure was preserved in speech materials. The main effect monetary incentive was not significant ($F(2, 36) = 2.23$; $p = 0.122$; $\eta^2_P = 0.11$), suggesting that varying monetary

incentive over 1, 3, and 5 cents did not significantly impact CNV amplitude. The fine structure × monetary incentive interaction approached statistical significance ($F(4, 72) = 2.23$; $p = 0.074$; $\eta^2_P = 0.11$). However, in contrast to the main effect of fine structure on CNV amplitude, the fine structure × monetary incentive interaction was not replicable at other fronto-central electrodes.

Second, we compared overall CNV amplitude and the impact of fine structure on CNV amplitude between main and follow-up experiment. To this end, we further collapsed average CNV amplitudes in the follow-up experiment across the three monetary incentive levels. Average CNV amplitudes were submitted to a repeated-measures ANOVA with the within-subject factor fine structure and the between-subject factor experiment. Critically, the main effect experiment was significant ($F(1, 35) = 30.54$; $p = 0.008$; $\eta^2_P = 0.19$), indicating that across experimental conditions, CNV magnitude was significantly larger (i.e., more negative) in the follow-up compared to the main experiment (Figure 3.9D). The fine structure × experiment interaction was not significant ($F(2, 70) = 0.24$; $p = 0.78$; $\eta^2_P < 0.01$), indicating a similar impact of fine structure on CNV amplitude in main and follow-up experiment.

3.2.4 Discussion

In the present follow-up experiment, we tested whether monetary incentives affect behavior and an elecrophysiological signature of selective attention (CNV) in the auditory number comparison task. Our results can be summarized as follows: (1) Varying monetary incentives across 1, 3, and 5 cent in the follow-up experiment did not significantly affect behavioral peformance or CNV magnitude. (2) Critically, however, overall (condition-independent) CNV magnitude under varying monetary incentives in the follow-up experiment was larger compared to the main experiment (chapter 3.1) where monetray incentives were not manipulated. (3) CNV magnitude increased with less fine structure in speech materials, replicating the central finding of the main experiment (chapter 3.1) in a different sample of participants.

3.2.4.1 No effect of monetary incentives in the follow-up experiment

In the follow-up experiment, monetary incentives varied over three levels (1, 3, and 5 cent) across trials. We found no significant effect of monetary incentives on performance (Figure 3.8) or CNV magnitude in the auditory number comparison task (Figure 3.9). This null-finding agrees with prior studies that also found no significant change in CNV magnitude when higher monetary incentives

were expected in case of correct performance (Goldstein et al., 2006; Broyd et al., 2012). There are different explanations why varying monetary incentives were ineffective in the follow-up experiment.

First, it might be that our participants' attentional effort was entirely unaffected by monetary incentives. Thus, higher incentives did not increase the attentional effort and consequently no modulation of behaviour or CNV magnitude was observed. However, we consider this rather unlikely, since prior studies have found beneficial effects of monetary incentives on performance in attention-demanding tasks (Small et al., 2005; Engelmann and Pessoa, 2007; Zedelius et al., 2012). Second, our findings could indicate that the CNV, as opposed to other ERP components (see Yeung and Sanfey, 2004; Sato et al., 2005), does not reflect participants' attentional effort. This way, it might be that participants' attentional effort was enhanced with higher monetary incentives. However, the CNV was insensitive to these changes in the deliberate allocation of attention to speech in noise. We consider a third possible explanation most plausible. Therefore, it might be that participants' attentional effort was already at maximum (ceiling) under the lowest incentive condition (1 cent), so that no further increase was possible under higher incentives (3 & 5 cent). The follow-up experiment was designed in a way to particularly emphasize monetary incentives. Thus, we provided feedback at the end of each trial to inform participants about whether they won or lost money on that trial. Furthermore, participants were instructed that they can keep the money they gain in the experiment and they were presented with their balance (i.e., sum over acquired and lost incentives over the experiment) after each trial. Thus, it is reasonable to assume that participants engaged a relatively high degree of attentional effort during the entire follow-up experiment, irrespective of the exact monetary incentives that were at stake on individual trials (see also below).

3.2.4.2 Enlarged CNV magnitude in the follow-up experiment

The most important finding of the present study resulted from the comparison of CNV magnitude in the main experiment (see also chapter 3.1) and in the follow-up experiment. We found that overall (condition-independent) CNV magnitude was larger in the follow-up compared to the main experiment (Figure 3.9). It is usually difficult to interpret the difference in overall ERP amplitude between experiments. This is because different laboratories typically use different hardware (e.g., EEG amplifiers, electrode systems) and data analysis software, which significantly affect the measured signal. Contrary, we used the same hard- and software for EEG recording and analysis in main and follow-up experiment (see Methods). Therefore, a direct comparison between CNV magnitudes in main and follow-up experiment was not problematic.

The major difference between main and follow-up experiment was the fact that monetary incentives were only manipulated in the latter. It is therefore reasonable to conclude that the presence of the manipulation of monetary incentives caused an increase in CNV magnitude (i.e., stronger scalp negativity). The direction of the observed effect agrees with our main hypothesis that monetary incentives should increase participants' attentional effort, which in turn should surface in an enlarged CNV magnitude.

Interstingly, we found no concomitant difference in behavioral measures between main and follow-up experiment (Figure 3.8). This is somewhat surprising since larger CNV magnitude is typically associated with improved performance in attention-demanding tasks (Rockstroh et al., 1993; O'Connell et al., 2009; see also chapter 3.1). Our findings might indicate that an increase in attentional effort does not necessarily improve detection of auditory signals. Thus, increasing attentional effort to select an auditory object (here: spoken digits) from noise can only be beneficial if the formation of the auditory object, which depends on acoustic stimulus properties, is successfully accomplished in the first place (Shinn-Cunningham and Best, 2008). In other words, selective attention does not improve performance if the to-be-attended signal is not properly defined.

It is important to note that also aspects of the experimental design other than the manipulation of monetary incentives might have driven the observed difference in CNV magnitude between main and follow-up experiment. Since monetary incentives were cued prior to the acoustic stimulation (Figure 3.7), the overall duration of single trials was longer in the follow-up experiment, which might have affected CNV magnitude. Moreover, participants received feedback concerning the accuracy of their performance only in the follow-up experiment. It might be that the presence of feedback allowed for a better performance-monitoring and enhanced participants' effort in the auditory number comparison task. This could imply that the follow-up experiment without a manipulation of monetary incentives but with performance feedback would have generated a similar increase in CNV magnitude. Future studies could investigate in detail in how far CNV magnitude is affected by trial duration, feedback, or monetary incentives.

3.2.4.3 CNV magnitude decreases with more fine structure in the follow-up experiment

With more preserved fine structure in speech materials, CNV magnitude in the follow-up experiment decreased (i.e., became less negative; Figure 3.9B). This is an important result since it replicates the major finding of the main experiment (chapter 3.1) in a different sample of young participants. Under

more degraded acoustic conditions, a larger CNV magnitude possibly reflects participants' enhanced allocation of attention to overcome listening challenges. The present finding demonstrates that this is a robust effect, which also holds when an additional experimental manipulation (i.e., varying monetary incentives) is added to the experimental design.

3.2.4.4 Conclusions

Our results suggest that the CNV is sensitive to participants' enhanced attentional effort in an experimental context where performance is related to monetary incentives. However, increasing monetary incentives from 1 to 5 cent do not further modulate CNV magnitude. Besides, our previous finding of enhanced CNV magnitude under more degraded acoustics (see main experiment in chapter 3.1) could be corroborated by the replication of this effect in a different study sample.

4 Study 2: Acoustics and predictions drive neural mechanisms of attention

This study describes two additional analyses of the same dataset of younger and older listeners investigated in Study 1.1. Study 2.1 analyses listeners' brain oscillatory dynamics in the auditory number comparison task. Study 2.2 investigates the phase-locking of the EEG signal to the acoustic envelopes of attended and unattended speech.

4.1 Study 2.1: Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits[4]

4.1.1 Introduction

Natural environments are rich of sensory information from both relevant (i.e., target) and irrelevant (i.e., noise) sources. Selective attention to relevant information enhances the neural representation of targets (Desimone and Duncan, 1995). According to the "functional inhibition" framework, neural alpha oscillations (~10 Hz) support target processing through the inhibition of task-irrelevant sensory modalities or brain processes (Jensen and Mazaheri, 2010). Alpha power is modulated by task-irrelevant sensory interference (Sauseng et al., 2009), by anticipation of distracting interference (e.g., Bonnefond and Jensen, 2012), and by predictions about non-distracting events (e.g., van Ede et al., 2011). The specific role of alpha oscillations in attentional processing is indicated, for example, by alpha power increase in parieto-occipital regions when attention shifts towards the auditory modality (Adrian, 1944; Foxe et al., 1998; Mazaheri et al., 2014).

Speech perception against competing talkers is a paradigmatic example for distracting interference but the role of alpha oscillations in these situations is weakly explored (Kerlin et al., 2010; Strauß et al., 2014). Alpha power increases if listening conditions become more demanding due to degradation of acoustic detail (Obleser et al., 2012; Becker et al., 2013), increasing syntactic complexity (Meyer et al., 2013), and low temporal expectancy (Wilsch et al., 2014). However, it is thus far unknown whether cues that allow listeners to predict upcoming information impinge upon alpha oscillations in a similar

[4] This section is adapted from the article published in the Journal of Neuroscience by Wöstmann, Herrmann, Wilsch, & Obleser (2015).

manner. Moreover, it is unclear whether listeners of different age utilize acoustic information and predictive cues similarly to overcome listening challenges in multi-talker situations.

For healthy older adults, listening in multi-talker situations is particularly effortful (Pichora-Fuller and Souza, 2003). This effort cannot be explained by sensory hearing loss alone (Wingfield et al., 2005; Meister et al., 2012). Instead, attentional control changes with age (Tun et al., 2002; Gazzaley et al., 2005a), which might involve that older listeners strongly attend to acoustic features of the stimulation (e.g., Passow et al., 2014) and that they are unable to ignore task-irrelevant acoustic stimuli (e.g., Chao and Knight, 1997). Importantly, age differences in neural responses might be driven by reduced sensory acuity in the elderly (Peelle et al., 2011) and therefore need rigorous experimental control. In the present study, stimulus intensities were individually adjusted for hearing acuity (frequency-specific adjustments to individual audiograms) and for speech-in-noise thresholds, in order to exclude "trivial" age effects in oscillatory alpha band dynamics related to decreasing stimulus audibility at an older age.

Participants performed a numerical comparison (Moyer and Landauer, 1967), where two spoken digits were embedded in a continuous stream of distracting speech. Acoustic detail (temporal fine structure; Moore, 2008b) and the degree to which the first digit predicted the second (Scheibe et al., 2010) varied orthogonally. Here we show that both stimulus manipulations (acoustic detail and predictiveness) modulate alpha power. We further provide evidence that these effects relate to subjective listening effort. Critically, aging affected behavioral performance as well as alpha power modulations by acoustic detail, suggesting that alpha power dynamics track age-related changes of listening behavior in challenging acoustic environments.

4.1.2 Materials and methods

4.1.2.1 Participants

Eighteen younger (mean age: 25.6 years; age range: 20–30; 9 females) and twenty older (mean age: 64.0 years; age range: 60–70; 11 females) healthy, right-handed German native speakers participated in the experiment. Data from two additional younger participants were recorded but excluded from the analysis due to technical problems during recording and overall below-chance task performance. Participants gave informed consent and were financially compensated for their participation. Procedures were in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Leipzig Medical faculty.

4.1.2.2 Hearing acuity

Participants' pure-tone air-conduction audiometric thresholds (at frequencies of 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz) were assessed by a trained audiologist separately for both ears in steps of 5 dB hearing level (HL) using a clinical audiometer (according to standardized procedures described in: BSA, 2011). Participants did not show interaural asymmetries (≥20 dB difference between both ears at more than two frequencies). Individual audiograms were used for frequency-specific adjustment of stimulus intensities (see below). Participants' audiograms, details concerning the individual stimulus adjustments, and the analysis of the event-related potential (ERP) for the same data set have been published before (Wöstmann et al., in press; see Study 1.1).

4.1.2.3 Number comparison task

Participants performed an auditory version of a number comparison task (Moyer and Landauer, 1967). In detail, each trial started with the visual presentation of the two response options ('kleiner', 'größer'; German for 'smaller' and 'larger', respectively) on the computer screen, followed (after 1.5 s) by the binaural presentation of a continuous speech masker and two sequentially presented spoken digits (Figure 4.1A). Following sound offset, participants indicated via button press on a response box whether the second digit was smaller (left button pressed with left thumb) or larger (right button pressed with right thumb) than the first. Subsequently, they rated their confidence in this decision on a three-point scale (1 = unconfident; 3 = confident). Participants were instructed to perform the number comparison as fast and as accurate as possible. The next trial started self-paced with an additional button press. To eliminate possible effects of participants' eye closure on alpha oscillations, participants were instructed to keep their eyes open during the trials. Participants were monitored via video camera to make sure that they did not close their eyes during acoustic stimulation. Stimulation was controlled by Presentation software (Neurobehavioral Systems).

4.1.2.4 Speech materials

German spoken digits ranging from 21 to 99 (excluding integer multiples of ten) were recorded from a trained female speaker (sampling rate, 44.1 kHz). All digits contained four syllables (mean digit length ± *SEM*: 1.125 ± 0.007 s). A distracting masker stimulus was extracted from a German audiobook (Oscar Wilde, "The young king", German title: "Der junge König") spoken by a different female talker (sampling rate, 44.1 kHz). To increase the energetic overlap of masker and digits, silent periods longer than 70 ms were removed automatically from the masker (using a customized Matlab script R2013a;

MathWorks Inc.). The resulting masker stimulus had a length of about 30 min, from which we extracted 1000 random snippets.

For each experimental stimulus, two different target digits (referred to as S1 and S2) and one masker snippet (referred to as masker) were selected randomly. Digits and masker were combined by adding the waveforms such that S1 and S2 were presented 0.5 s and 3.125 s after masker onset, respectively. The interval between S1 offset and S2 onset was on average 1.5 s (depending on S1 duration). Stimuli ended with S2 offset and had an average duration of about 4.25 s (Figure 4.1A).

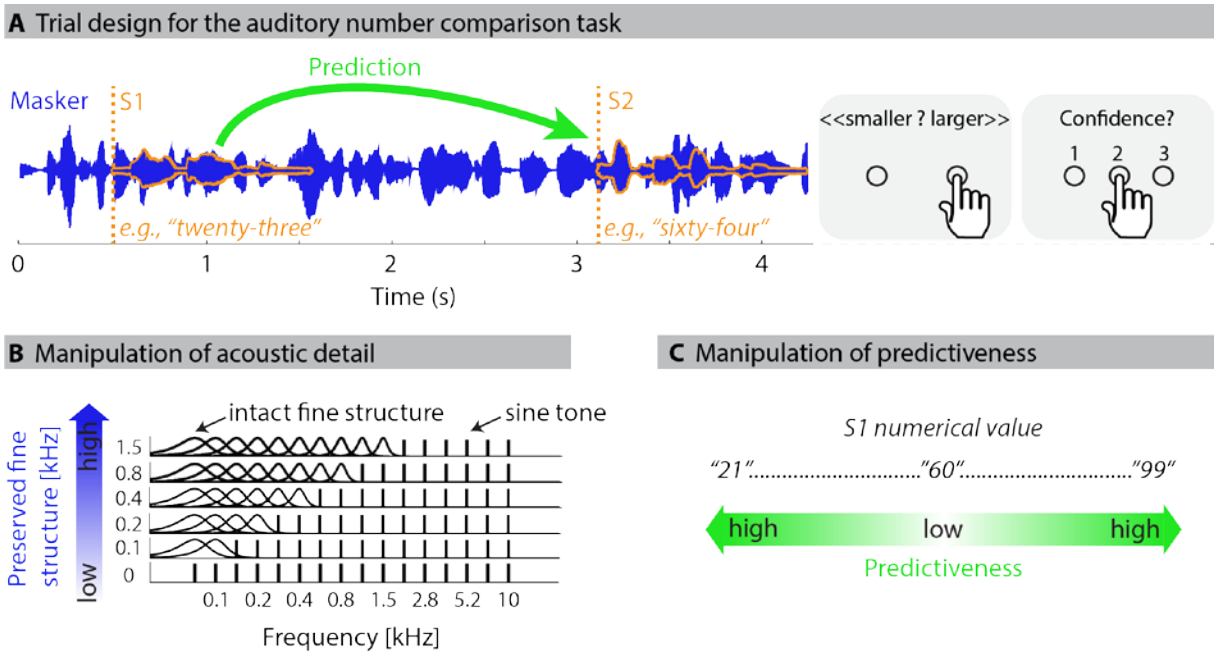

Figure 4.1. Trial design and experimental manipulation. (A) On each trial, participants listened to two spoken digits (S1 & S2, orange) embedded in a distracting speech masker. Their task was to indicate whether S2 was smaller or larger than S1, and how confident they were in this decision. (B) Acoustic detail was varied over six levels by parametrically preserving temporal fine structure of the signal's low frequencies (blue color gradient). (C) The degree to which S1 was predictive of the numerical value of S2 was operationalized as the numerical distance between S1 and the midpoint of all possible numbers (60; green color gradient).

4.1.2.5 Experimental conditions

In the current study, stimuli were manipulated along two orthogonal dimensions: acoustic detail and predictiveness. For the acoustic detail manipulation, the temporal fine structure of the combined signal (composed of masker and digits) was manipulated. In detail, the signal was divided into 16 overlapping frequency channels using a gammatone filterbank (implemented in the auditory toolbox for Matlab; Slaney, 1993). Channel center frequencies increased exponentially from 0.08 to 10 kHz. Six temporal fine structure conditions comprising different levels of acoustic degradation were generated. For a particular condition, frequency channels above one of six fine structure cut-offs (0, 0.11, 0.21, 0.4, 0.76, and 1.45 kHz) were degraded, while channels below and including the cut-off were left

unchanged (Hopkins et al., 2008; Figure 4.1B). In channels above the cut-off, the speech envelope was extracted using the Hilbert transform (Smith et al., 2002). The envelope was used to modulate a sinusoidal tone with random starting phase at the channel's center frequency. The resulting signal was filtered again with the initial gammatone filters to remove out-of-channel frequency components (Lunner et al., 2012). Finally, intact and modified channels were combined, yielding six different levels of temporal fine structure preservation. In sum, this manipulation of the temporal fine structure degraded fast spectro-temporal fluctuations, rendering the perceptual segregation of digits and masker more demanding. Slow temporal envelope fluctuations were left intact (Shamma and Lorenzi, 2013). Critically, degraded stimuli were intelligible as the number of frequency channels (16) was sufficiently high (Shannon et al., 1995; Obleser et al., 2007; Obleser et al., 2008).

For the predictiveness manipulation, the degree to which the S1 digit was predictive of the S2 digit was operationalized as the numerical distance between S1 and the midpoint of all possible digits. In detail, digits in the experiment ranged between 21 and 99, meaning that 60 was the midpoint of all digits. When the S1 digit was considerably smaller than 60, participants could predict that the S2 digit would likely be larger than S1, and vice versa for S1 digits larger than 60. Contrary, if the S1 digit was close to 60, no prediction about whether S2 would be smaller or larger could be made. Thus, with increasing numerical distance between S1 and 60, participants could better predict whether S2 would be smaller or larger (Figure 4.1C).

4.1.2.6 Individual stimulus adjustments

Prior to the actual experiment, stimuli underwent a frequency-specific amplification (CAMEQ; Moore et al., 1998) to account for considerable differences in hearing thresholds estimated in the audiograms, especially between age groups. This procedure aimed at the same overall perceived stimulus loudness for all participants corresponding to a stimulus intensity of ~75 dB SPL for a listener with average normal hearing (audiometric thresholds of 0 dB HL at all test frequencies).

Since speech-in-noise hearing thresholds cannot be matched between age groups by controlling only for pure-tone audiometric thresholds (see Pichora-Fuller et al., 1995) an additional adaptive tracking procedure (Levitt, 1971) was used to estimate the digit-to-masker sound-level ratio yielding 70.9% correct responses in our number comparison task under the most extreme acoustic degradation (0 kHz fine structure cut-off). To this end, the sound level of the digits was adapted while keeping the masker sound level fixed at –30 dB full-scale (RMS – root mean square). Mean digit-to-masker sound-

level ratio for young participants was –20.83 dB (±0.72 SEM) and for older participants –15.35 dB (±1.29 SEM), which were significantly different ($t_{36} = 3.60$; $p = 0.001$; $r = 0.51$).

4.1.2.7 Procedure

After the individual stimulus adjustments were applied, participants took part in the main experiment. Acoustic stimulation and EEG recording were carried out in an electrically-shielded and sound-attenuated booth. Participants were seated in a comfortable chair in front of a computer screen. Auditory stimuli were presented via TDH39 audiometric headphones. Each participant performed 300 trials, 50 for each temporal fine structure cut-off level. The experiment was divided into five blocks. Each block contained ten trials of each fine structure cut-off in random order. Predictiveness of the second digit was fully randomized across the 300 trials. That is, the numerical values of S1 and S2 varied randomly across trials with the constraint that in half of the trials S2 was larger than S1 and in the other half smaller than S1 (S1 and S2 digits were never equal). The experiment lasted approximately 70 minutes.

4.1.2.8 Statistical analysis of behavioral data

Participants' performance in the auditory number comparison task was quantified using weighted percentage correct responses (weighted accuracy). In detail, the binary response in each trial (correct vs incorrect) was weighted by the trial's confidence rating to get a more fine grained (six-level) measure of task performance (Kitayama, 1991; Herrmann et al., 2014). To this end, a correct response was transformed to 100% weighted accuracy in case of a high confidence rating, to 80% in case of medium confidence, and to 60% in case of low confidence. Similarly, an incorrect response was transformed to 40% weighted accuracy for a low confidence rating, to 20% for medium confidence, and to 0% for high confidence. In the remainder of this paper we use, for simplicity, the term 'accuracy' to refer to accuracy weighted by confidence ratings.

As a second performance measure, we quantified participants' response times in the number comparison task. Response times corresponded to the time interval between the onset of the second digit and participants' button press to indicate whether the second digit was smaller or larger than the first.

Changes in behavioral performance (accuracy and response times) as a function of acoustic detail were tested as follows: For each temporal fine structure cut-off level, single-trial accuracy values and response times were averaged (ignoring predictiveness). For each participant, a linear function was

fitted to the averaged accuracy values and response times as a function of six linearly spaced fine structure cut-offs (predictor values: –2.5, –1.5, –0.5, 0.5, 1.5, 2.5). Linear spacing of cut-offs for fitting was used because logarithmic spacing of physical stimulus frequencies relates to linear spacing in auditory perception (Attneave and Olson, 1971). The estimated linear coefficients were subsequently tested against zero using one sample t-tests. Significant differences from zero would indicate modulation of behavioral performance by acoustic detail.

Changes in behavioral performance (accuracy and response times) as a function of predictiveness were examined as follows: The degree to which S1 was predictive of S2 on each trial was quantified as the absolute numerical difference between S1 and 60 (average digit across the experiment). Predictiveness values across all trials were divided into six percentile bins (no overlap), and single-trial accuracy values and response times were averaged within each bin. Linear functions were fitted to the averaged accuracy values and response times for each participant as a function of percentile bins (zero-centered predictor values: –2.5, –1.5, –0.5, 0.5, 1.5, 2.5). The estimated linear coefficients were tested against zero using one sample t-tests. Significant differences from zero would indicate modulation of behavioral performance by predictiveness.

In order to test for differences between performance modulation by acoustic detail and predictiveness as well as for differences between age groups, repeated-measures ANOVAs were carried out. The within-subject factor was stimulus dimension (acoustic detail vs predictiveness) and the between-subject factor age group (younger vs older) using the estimated linear coefficients for accuracy and response times as dependent measures. To follow up on significant age group × stimulus dimension interactions, post-hoc independent samples t-tests were used to test for effects of age group on linear coefficients separately for the manipulation of acoustic detail and predictiveness.

Overall performance between age groups was compared by submitting participants' average accuracy and average response times (across all manipulation levels) to independent samples t-tests.

4.1.2.9 Electroencephalography (EEG) recording

Electroencephalograms were recorded at a 500-Hz sampling rate with a DC–135 Hz filter pass band (TMS international, Enschede, The Netherlands). Twenty-eight electrodes (Ag/Ag-Cl) were placed at the following positions (Easycap, Herrsching, Germany): Fpz, Fp1, Fp2, Fz, F3, F4, F7, F8, FC3, FC4, FT7, FT8, Cz, C3, C4, T7, T8, CP5, CP6, Pz, P3, P4, P7, P8, O1, O2, left mastoid (A1), and right mastoid (A2). The reference electrode was placed at the tip of the nose and the ground electrode at the

sternum. The electrooculogram was recorded from vertical and horizontal bipolar montages. All electrode resistances were kept below 5 kΩ.

Data were analyzed offline using custom Matlab scripts and the Fieldtrip toolbox (Version 2013-01-14; Oostenveld et al., 2011). Epochs were extracted from the continuous signal time-locked to masker onset (–1.5 to 5.5 s). Epochs were low-pass filtered at 100 Hz. An independent components analysis (ICA) was performed on the epoched data. Components corresponding to eye blinks, saccadic eye movements, muscle activity, electrode drifts, and heartbeats were identified and rejected by inspection of the components' topographies, frequency spectra, and time courses. Remaining artifact-contaminated trials were deleted after visual inspection of EEG waveforms at all electrodes. On average, 7 ± 1 % (*SEM*) trials in each participant were rejected from further analyses.

Time–frequency representations of single trials were estimated by convolving the single-trial time series with a family of Morlet wavelets between 1 and 30 Hz (in steps of 0.5 Hz; width: 7 cycles) and from –1.5 to 5.5 s (in steps 0.02 s). Single-trial power was obtained by squaring the magnitude of the estimated complex wavelet transform coefficients. Power changes relative to a pre-stimulus baseline were computed by means of subtraction and division by the average power from –0.8 to 0 s (relative change baseline).

4.1.2.10 Overall temporal dynamics of alpha power

We analyzed the overall time course of alpha power during the number comparison task, irrelevant of varying acoustic detail and stimulus predictiveness (Figure 4.3). To this end, single-trial oscillatory power was averaged across all conditions, frequency bins in the alpha band (7–13 Hz), and five parietal electrodes exhibiting the strongest alpha power (CP5, P3, Pz, P4, CP6; Figure 4.3A). To test for age effects, time courses of alpha power were compared between age groups by contrasting average alpha power estimates in steps of 0.02 s with independent samples t-tests. P-values from multiple t-tests were adjusted to control the false discovery rate (FDR; Benjamini and Hochberg, 1995).

4.1.2.11 Effects of experimental manipulations on EEG data

Modulatory influences of acoustic detail and predictiveness on oscillatory power were analyzed as follows. For each participant, two linear functions were fitted to single-trial power values (independently for each time-frequency bin and electrode), first, as a function of acoustic detail and second, as a function of predictiveness percentile bins (using parametric regression t-tests for independent samples implemented in the *ft_freqstatistics* function in Fieldtrip; predictor values: –2.5, –

1.5, –0.5, 0.5, 1.5, 2.5). This resulted in one time–frequency–electrode matrix of estimated linear coefficients for the acoustic detail manipulation and in one matrix for the predictiveness manipulation, reflecting the modulation of single-trial power for each participant.

For the statistical analysis across participants, we focused on the alpha frequency range (~10 Hz) for which we hypothesized to observe power changes due to manipulations of acoustic detail and predictiveness (see *Introduction*). Furthermore, analyses were conducted including participants of both age groups, followed by analyses of age differences where effects for all participants (younger and older) were significant. To this end, estimated linear coefficients in the 7–13 Hz frequency band, the 0–5.2 s time window, and all scalp electrodes were tested against zero using two cluster-based permutation one sample t-tests (Maris and Oostenveld, 2007), one for the effect of acoustic detail and one for the effect of predicitveness. These tests clustered t-values of adjacent bins in time–frequency–electrode space with a p-value smaller than 0.05, considering a minimum of three neighboring electrodes as a cluster. The summed t-value of each cluster was computed and compared against the distribution of 1000 iteratively and randomly drawn clusters from data for which condition labels were permuted. The cluster p-value resulted from the proportion of Monte Carlo iterations in which the summed t-statistic of the observed cluster was exceeded. As we performed this analysis as a two-sided test (for clusters exhibiting positive and negative effects), clusters with $p < 0.025$ were considered significant. This analysis revealed four significant clusters, two for the effect of acoustic detail and two for the effect of predictiveness (Figure 4.4).

In order to test for an effect of age group on linear coefficients in all four clusters, a repeated-measures ANOVA (within-subject factor: cluster; between-subject factor: age group) was calculated for the averaged linear coefficients in the four clusters. Since task-related power suppression is known to depend on overall power (Doppelmayr et al., 1998; Klimesch et al., 2003), we controlled for effects of overall alpha power and the decrease in alpha power over the trial time course (Figure 4.3) in two additional ANOVAs: For the first additional ANOVA, we extracted overall alpha power (averaged across conditions) at those time–frequency–electrode bins of the four significant clusters, resulting in four covariates that were included in the repeated-measures ANOVA. For the second additional ANOVA, linear coefficients estimated from linear fits to the overall alpha power (averaged across 7–13 Hz and electrodes CP5, P3, Pz, P4, CP6) as a function of time (ranging from 0.82 s to 4.88 s, that is from cluster A1 to A2) were included as a covariate. To follow up a significant age group × cluster

interaction, post-hoc independent samples t-tests were used to test for effects of age group in each cluster.

To test for a possible interaction between acoustic detail and predictiveness on alpha power modulation, we averaged alpha power estimates of only those time–frequency–electrode power bins that had been part of both the acoustic detail and the predictiveness clusters in the analyses outlined above (Obleser et al., 2012). These average power estimates were submitted to a repeated-measures ANOVA (within-subject factors: acoustic detail, predictiveness; between-subject factor: age group).

4.1.2.12 Effect sizes

To estimate effect sizes for F-statistics (ANOVAs), we calculated the partial eta-squared (η^2_P). For t-statistics (dependent and independent samples t-tests), we estimated the effect size measure r, which is bound between 0 and 1 (Rosenthal, 1994). Effect sizes for multiple t-tests (e.g., for all time–frequency–electrode bins belonging to a significant cluster) were estimated by averaging r values across individual tests into a composite cluster-effect size R.

4.1.2.13 The relation between alpha oscillations and subjective difficulty measures

We further tested whether alpha power modulations (by acoustic detail and predictiveness) within observed clusters were related to participants' subjective listening effort in background noise and confidence ratings. In detail, alpha power modulation was quantified as the average of the linear coefficients across significant time points, frequency bins, and electrodes as well as across the four significant clusters of acoustic detail and predictiveness. Subjective listening effort in background noise was quantified as the response in a post experiment inquiry where participants answered the question *'In general, how difficult is it for you to listen to a single speaker if several other people are talking loudly in the background?'* on a 5-point Likert scale (1 = easy, 5 = difficult; question translated from German). The Spearman correlation was obtained between ratings of subjective listening effort in noise and alpha power modulation (Figure 4.5A).

To test for an impact of alpha power on participants' subjectively experienced certainty in the numerical decision, we analyzed whether alpha power in trials with the same level of acoustic detail and predictiveness would affect participants' confidence ratings. To this end, we averaged participants' single-trial alpha power (7–13 Hz) across all scalp electrodes and in the time period of significant alpha power modulations, that is, between the onset of the earliest significant cluster (0.82 s) and the offset of the latest significant cluster (4.88 s). Subsequently, within each combination of the 6 (acoustic detail) ×

6 (predictiveness) levels, trials were divided into three (non-overlapping) percentile bins based on alpha power. In detail, trials were assigned to low, medium, or high alpha power, depending on whether their alpha power was amongst the lowest third (0–33%), medium third (33–66%), or highest third (66–100%) out of all trials within a particular acoustic detail–predictiveness combination. Next, the average over the trial's confidence ratings in these three percentiles was calculated. Critically, the sorting of single trials according to alpha power was carried out independently for each combination of acoustic detail and predictiveness, and average confidence ratings were thus independent of between-condition effects. For each participant, mean confidence ratings were subsequently averaged over all combinations of acoustic detail and predictiveness levels, to obtain three confidence values for low, medium, and high alpha power trials, respectively. For each participant, a linear function was fitted to confidence values as a function of alpha power bin (predictor values: –1, 0, 1). Estimated linear coefficients across participants were tested against zero using a one sample t-test. A significant difference from zero would indicate a modulation of confidence ratings by alpha power. Linear coefficients of younger and older participants were compared using an independent samples t-test (Figure 4.5B).

4.1.3 Results

4.1.3.1 Acoustic detail and predictiveness enhance performance

Figure 4.2A shows mean accuracy and response times in the number comparison task as a function of acoustic detail and predictiveness. Parametric variations along either acoustic detail or predictiveness were quantified as the estimated coefficient from linear fits to accuracy and response times. Testing the linear coefficients against zero revealed that, with higher levels of acoustic detail, accuracy increased ($t_{37} = 17.81$; $p < 0.001$; $r = 0.95$) and response time decreased ($t_{37} = -6.95$; $p < 0.001$; $r = 0.75$). Similarly, for higher levels of predictiveness, accuracy increased ($t_{37} = 5.92$; $p < 0.001$; $r = 0.70$) and response time decreased ($t_{37} = -5.31$; $p < 0.001$; $r = 0.66$). The difference in overall accuracy between age groups approached significance ($t_{36} = 1.95$; $p = 0.059$; $r = 0.31$), indicating a slightly higher overall task accuracy for older participants. Overall response times relative to S2 onset did not differ between age groups (mean response time younger: 2.28 s, older: 2.23 s; $t_{36} = 0.28$; $p = 0.783$; $r = 0.05$).

Differential effects of stimulus dimension (acoustic detail vs predictiveness) and age group (younger vs older) on linear coefficients for accuracy and response times were analyzed using repeated-measures ANOVAs (within-subject factor: stimulus dimension; between-subject factor: age

group). The main effect of stimulus dimension was significant for accuracy ($F(1, 36) = 117.88$; $p < 0.001$, $\eta^2_P = 0.77$) and for response times ($F(1, 36) = 22.38$; $p < 0.001$, $\eta^2_P = 0.38$), showing that the modulation of performance (i.e., accuracy increase and response time decrease) was stronger for acoustic detail than for predictiveness (Figure 4.2B). There was no significant main effect of age group on linear coefficients for accuracy ($F(1, 36) = 0.14$; $p = 0.714$, $\eta^2_P = 0.004$) but on linear coefficients for response times ($F(1, 36) = 4.52$; $p = 0.040$, $\eta^2_P = 0.11$), indicating that the decrease in response times was stronger in older than younger participants, regardless of the stimulus dimension.

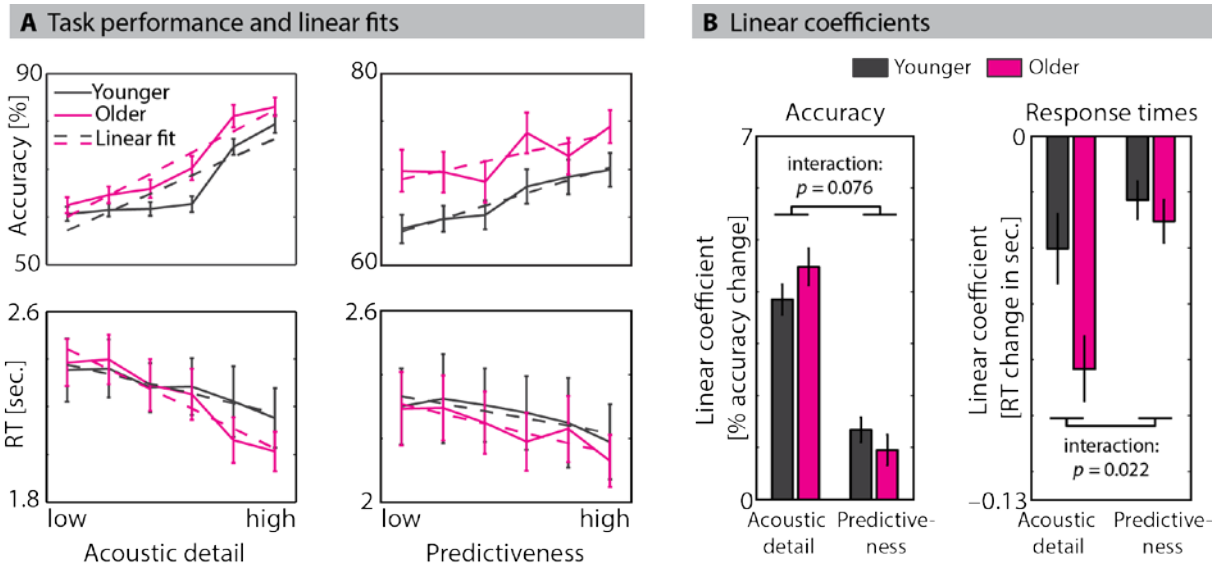

Figure 4.2. Effects of acoustic detail and predictiveness on task performance. (A) Mean accuracy (top row) and response times (RT; bottom row) as a function of acoustic detail (left column) and predictiveness (right column). Solid lines show average accuracy and response times for younger (black) and older (magenta) participants. Dashed lines show the average of linear fits to accuracy and response times along parametric variations of acoustic detail and predictiveness. Note the different scaling of y-axes. (B) Bars indicate average linear coefficients quantifying the increase in accuracy and decrease in response times with each level of acoustic detail or predictiveness for younger (black) and older (magenta) participants. The age group × stimulus dimension (acoustic detail vs predictiveness) interaction approached significance for accuracy measures (p = 0.076) and reached significance for response times (p = 0.022). Error bars indicate ± 1 SEM.

Critically, the age group × stimulus dimension interaction on linear coefficients for accuracy approached statistical significance ($F(1, 36) = 3.34$; $p = 0.076$; $\eta^2_P = 0.09$) and reached statistical significance for linear coefficients for response times ($F(1, 36) = 5.69$; $p = 0.022$; $\eta^2_P = 0.14$). Figure 4.2B indicates the direction of these interactions. Older compared with younger participants' accuracy and response times were affected more strongly with more acoustic detail, whereas predictiveness diminished (for response times) or reversed (for accuracy) this age difference. Post-hoc tests for age effects revealed that the linear coefficients quantifying changes in response times with acoustic detail were significantly smaller for older compared to younger participants ($t_{36} = 2.53$; $p = 0.016$; $r = 0.39$). All remaining pairwise comparisons did not reach statistical significance (all $p > 0.15$; all $r < 0.22$).

4.1.3.2 *Temporal dynamics of alpha oscillations*

Before testing effects of varying acoustic detail and stimulus predictiveness on alpha power, we analyzed temporal dynamics of overall alpha power (7–13 Hz) across all trials, that is, independent of experimental conditions. Alpha power was enhanced during acoustic stimulation (0–4.25 s), before returning to baseline at the end of the trial (Figure 4.3A). The increase in alpha power was strongest at parietal electrode sites. We tested whether the time course of overall alpha power (averaged across five parietal electrodes) differed between age groups (Figure 4.3B). Alpha power was lower in older than younger participants only towards the end of the trial (> 3.9 s; multiple independent samples t-tests for 20 ms time intervals; $p < 0.05$, FDR-corrected; $R = 0.51$).

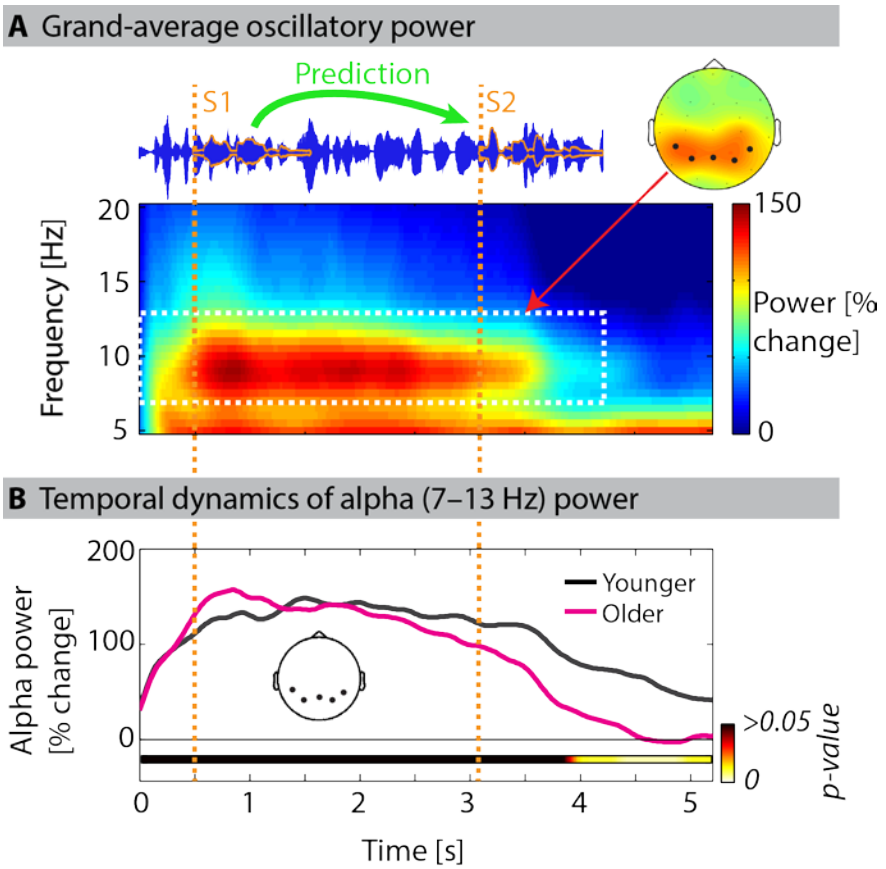

Figure 4.3. Overall temporal dynamics of alpha power. (A) Grand average overall oscillatory power (averaged across both age groups and all scalp electrodes) during the auditory number comparison task. Topographical map shows alpha power (7–13 Hz) during acoustic stimulation (0–4.25 s). Electrodes exhibiting strongest alpha power are highlighted. (B) Time courses of mean alpha power (averaged across five parietal electrodes) for younger (black) and older (magenta) participants. *P*-values for the comparison of alpha power between age groups (multiple independent samples t-tests for 20 ms time intervals; FDR-corrected) in horizontal bar indicate lower alpha power for older participants in the end of the trial (> 3.9 s).

4.1.3.3 *Acoustic detail and predictiveness modulate alpha power*

Figure 4.4A shows the effects of increasing acoustic detail and predictiveness on alpha power (7–13 Hz). Both effects were quantified by linear coefficients (slopes) reflecting the change in alpha power

with increasing levels of acoustic detail or with increasing levels of predictiveness. Cluster-based permutation tests revealed two significant clusters for the effect of acoustic detail, referred to as 'A1' (p = 0.006; R = 0.41; 0.82–1.92 s) and 'A2' (p < 0.001; R = 0.46; 3.52–4.88 s). Similarly, two significant clusters were found for the effect of predictiveness, referred to as 'P1' (p < 0.001; R = 0.41; 1.22–2.60 s) and 'P2' (p = 0.015; R = 0.39; 2.64–3.68 s). No significant positive clusters were observed. For the significant negative clusters, linear coefficients were significantly smaller than zero. That is, alpha power decreased with increasing acoustic detail (higher temporal fine structure cut-offs) and increasing predictiveness (higher absolute numerical difference between S1 and 60). For all significant clusters these alpha power decreases were significant at a large number of electrodes (Figure 4.4B, topographic maps): Clusters A1, A2, and P1 spanned 26 of 28 scalp electrodes; cluster P2 spanned 20 of 28 scalp electrodes.

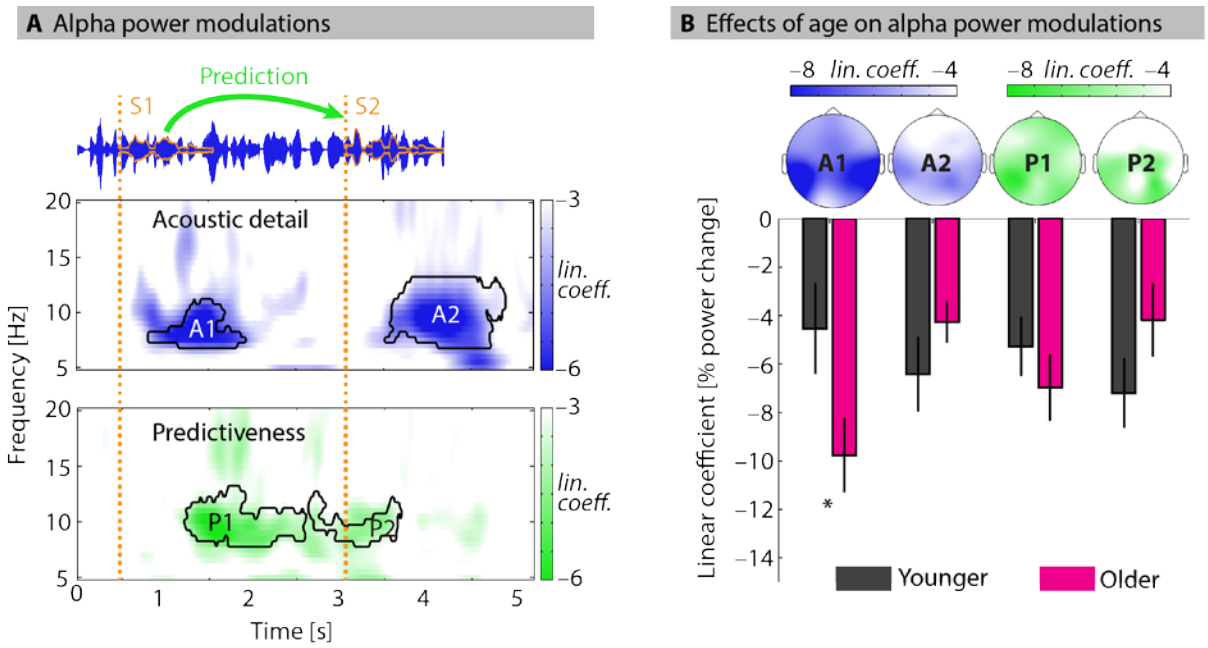

Figure 4.4. Effects of acoustic detail and predictiveness on alpha power. (A) Alpha power (7–13 Hz) decreased significantly with higher levels of acoustic detail (top panel; clusters A1 & A2) and higher levels of predictiveness (bottom panel; clusters P1 & P2) in distinct time periods. Estimated linear coefficients indicate the relative change in alpha power (in %) with each level of acoustic detail or predictiveness. (B) Topographical maps of clusters show a global decrease of alpha power with acoustic detail and predictiveness with the largest power decrease over centro-parietal electrode sites. Bars indicate average linear coefficients for younger (black) and older (magenta) participants. Significantly smaller linear coefficients for older participants in cluster A1 indicated a stronger alpha power modulation as a function of acoustic detail for older listeners around S1 offset (p = 0.036, uncorrected). Error bars indicate ± 1 *SEM*.

Critically, the temporal occurrence of significant clusters matched precisely with the manipulations of acoustic detail and predictiveness. That is, alpha power decreased during and shortly after spoken digits (S1 and S2) when more acoustic detail facilitated the encoding of digits (clusters A1 and A2). When S1 was better predictive of S2, alpha power decreased significantly during the time period when S2 could be predicted, that is, between the presentation of S1 and S2 (clusters P1 and P2).

In order to test whether the effects of acoustic detail and predictiveness on alpha power interact, power estimates for the conjunction of clusters in time–frequency–electrode space were submitted to repeated-measures ANOVAs (factors: acoustic detail, predictiveness). For the two conjunctions of clusters in the present data (A1 ∩ P1, A2 ∩ P2) neither the two-way interaction acoustic detail × predictiveness, nor the three-way interaction with age group was significant (all $p > 0.3$; all $\eta^2_P \leq 0.03$), thus indicating independent influences of acoustic detail and predictiveness on alpha power.

The effect of age group on linear coefficients in the four significant clusters was tested with a repeated-measures ANOVA (within-subject factor: cluster; between-subject factor: age group). Modeling the four clusters within one factor acknowledges that the clusters were temporally independent as they occurred in distinct (only partly overlapping) time intervals. The main effects of cluster ($F(3, 108) = 1.13$; $p = 0.34$; $\eta^2_P = 0.03$; no significant violation of sphericity: Mauchly's test, $p = 0.26$) and age group ($F(1, 36) = 0.08$; $p = 0.782$; $\eta^2_P < 0.01$) were not significant. However, the age group × cluster interaction was significant ($F(3, 108) = 6.58$; $p < 0.001$; $\eta^2_P = 0.16$), indicating a different pattern of alpha power modulations in the four clusters for younger compared with older participants (Figure 4.4B; note that this interaction was also significant using overall alpha power in the four clusters and the linear decrease in overall alpha power during the trial as covariates; $p = 0.047$ and $p < 0.001$, respectively). Post-hoc tests for differences between age groups in the four clusters revealed a significant effect of age group on linear coefficients in the A1 cluster ($t_{36} = 2.17$; $p = 0.036$; $r = 0.34$, uncorrected), but not in the three remaining clusters (A2, P1, P2; all $p > 0.15$; all $r < 0.24$). That is, older participants' alpha power during the encoding of S1 decreased stronger with increasing acoustic detail compared with younger participants.

Corroborating this age difference, the significant A1 cluster was found only for the group of older participants when the cluster analysis was performed separately for the two age groups. Note that this finding was well in line with behavioral results (Figure 4.2) where varying acoustic detail also had a relative stronger impact on older participants' task performance.

4.1.3.4 Alpha oscillations predict subjective measures of difficulty

An important question of the present study was whether participants' self-rated difficulty of speech-in-noise listening and their confidence in the numerical comparison were related to fluctuations in alpha power. We investigated this question with respect to subjective ratings of listening effort (self-rated after the experiment) and confidence ratings in the end of each experimental trial.

The mean estimated linear coefficients (reflecting modulation of alpha power by acoustic detail and predictiveness) across all clusters (Figure 4.4) were correlated with participants' self-reported effort of listening to a single speaker in the presence of background noise (Figure 4.5A). We observed a significant positive correlation ($r = 0.484$; $p = 0.002$; $df = 36$), indicating that listeners who experienced higher subjective listening effort exhibited weaker alpha power modulations with varying acoustic detail and predictiveness. The correlation was also significant when calculated for both age groups separately (younger: $r = 0.54$; $p = 0.021$; $df = 16$; older: $r = 0.48$; $p = 0.032$; $df = 18$).

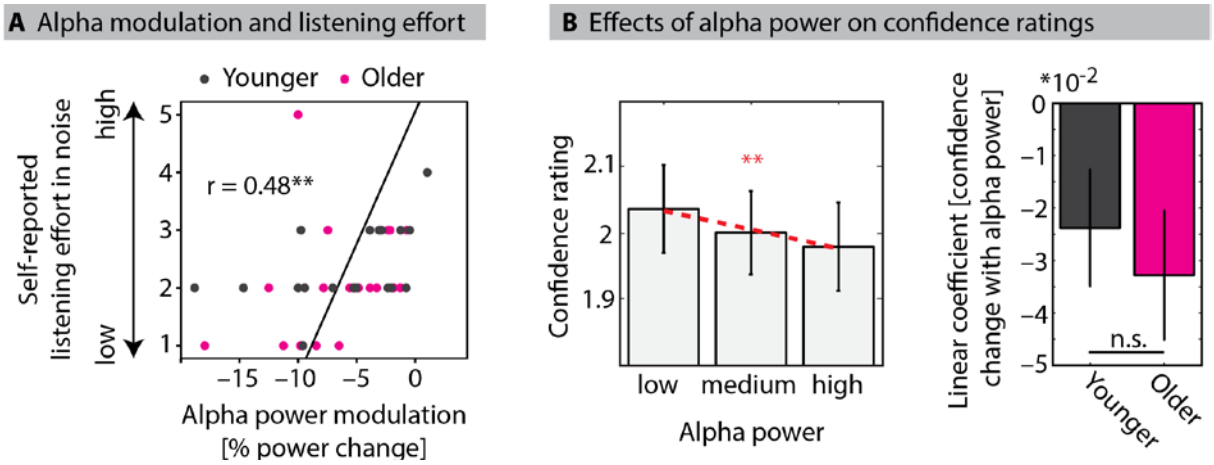

Figure 4.5. Alpha power relates to subjective difficulty measures. (A) Alpha power modulation (reflecting mean linear coefficients of alpha power changes with each level of acoustic detail and predictiveness across four significant clusters) as a function of participants' self-reported subjective listening effort in background noise (significant Spearman correlation, p = 0.002). Younger and older participants' data are shown in black and magenta, respectively. (B, left panel) Bars indicate mean confidence ratings for trials with low, medium, and high alpha power. Between-condition effects were eliminated by binning trials according to alpha power separately for each factor combination of acoustic detail and predictiveness. The mean linear fit to confidence ratings is indicated by the red dashed line. Confidence ratings decreased with higher alpha power (p = 0.001). (B, right panel) Bars indicate the average linear coefficients (quantifying changes in confidence ratings with each level of alpha power) for younger (black) and older (magenta) participants. Error bars indicate ± 1 SEM. **p < 0.01, n.s. not significant.

We tested whether alpha power would correlate with confidence ratings independent of variations in acoustic detail and predictiveness. To this end, mean alpha power (7–13 Hz) for single trials in the time period from the onset of the earliest significant cluster (A1, 0.82 s) until the offset of the latest significant cluster (A2, 4.88 s) at all scalp electrodes was determined. Single-trial confidence ratings within each factor combination in the 6 (acoustic detail) × 6 (predictiveness) design were binned into three alpha power percentiles (no overlap; low, medium, and high alpha power). Next, single-trial confidence ratings were averaged within each bin, and then across the 6 × 6 levels, resulting in three values for each participant, reflecting confidence in trials with low, medium, and high alpha power (Figure 4.5B). The coefficients from linear fits to changes in confidence ratings over these three levels of alpha power were significantly smaller than zero ($t_{37} = -3.44$; $p = 0.001$; $r = 0.50$). That is, confidence

ratings were higher in trials with lower alpha power. Linear coefficients did not differ significantly between age groups ($t_{36} = 0.54$; $p = 0.59$; $r = 0.09$; Figure 4.5B, right panel).

4.1.4 Discussion

We tested whether alpha oscillations track changing task demands in a multi-talker situation in younger and older listeners. Results can be summarized as follows: (1) Alpha power decreased with increasing acoustic detail and, critically, also with increasing stimulus predictiveness. (2) In older participants, increased acoustic detail induced a stronger behavioral benefit and a stronger alpha power decrease. (3) Stronger alpha power modulations with acoustic detail and predictiveness, as well as lower overall alpha power predicted lower subjective difficulty.

4.1.4.1 Listening demands modulate alpha oscillatory power

Behavioral results show that accuracy in a two-talker auditory number comparison task increased with more acoustic detail (temporal fine structure) in the stimulus materials and also with better numerical predictiveness (Figure 4.2). This agrees with previous research showing first, that preserved temporal fine structure facilitates perceptual segregation of competing talkers (e.g., Hopkins et al., 2008; Hopkins and Moore, 2010; Lunner et al., 2012) and second, that numerical predictiveness improves stimulus comparison (Scheibe et al., 2010). We extend previous observations by relating manipulations of acoustic detail and predictiveness to neural alpha oscillations.

On the neurophysiological level, alpha power decreased in distinct time intervals with parametric variations along two stimulus dimensions: First, with increasing acoustic detail, alpha power decreased during the encoding of target digits (Figure 4.4). This is consistent with previous observations of reduced alpha power for less degraded speech materials (e.g., Obleser and Weisz, 2012; Obleser et al., 2012). Although acoustic detail was manipulated during the entire trial, alpha power modulation occurred exclusively during the encoding of task-relevant digits. This suggests that the modulation of alpha power with acoustic detail is guided by attention to target signals.

Second, with better stimulus predictiveness, alpha power decreased during the prediction of the second digit (i.e., between the two digits). Although alpha power modulations have been found for varying temporal predictions of 'when' a target stimulus would occur (e.g., Rohenkohl and Nobre, 2011; Wilsch et al., 2014), evidence for the prediction of 'what' the target stimulus will be have so far been rare (for review, see Arnal and Giraud, 2012). Thus, in the current study we show that alpha power modulations reflect the predictiveness of upcoming semantic content. Stimulus predictiveness is

a complementary source of information (separate from acoustic detail) that listeners can use to reduce the uncertainty in the numerical comparison. In highly predictive trials, participants gather information in favor of a 'smaller' or 'larger' decision already with the first digit. Thus, increasing decision certainty surfaced as a relative reduction in alpha power and reduced listening demands.

Good performance in our number comparison task required selective attention to digits while ignoring the irrelevant speech masker ("cocktail party problem", Cherry, 1953). Enhanced alpha power at parieto-occipital sites when attention is directed towards the auditory modality is an established observation (Adrian, 1944; Foxe et al., 1998; Mazaheri et al., 2014). Based on previous localizations of alpha power effects in auditory tasks (Obleser and Weisz, 2012; Obleser et al., 2012), the current parietal distributions of alpha power likely originate from parietal cortex, which belongs to the "dorsal attention network" (Sadaghiani et al., 2010). Increased task difficulty (less acoustic detail or predictiveness) requires more attention to the auditory sensory input. Thus, task-irrelevant sensory modalities (e.g., vision) and task-irrelevant brain processes might be inhibited. Inhibition is likely reflected by enhanced alpha oscillations in a parietal network, which interacts with sensory areas during attention (Banerjee et al., 2011).

4.1.4.2 Age-related changes in listening behavior and alpha power dynamics

Overall alpha power was prominently enhanced during the number comparison task (see also Spitzer et al., 2014), but was reduced towards the end of a trial in older participants (Figure 4.3B). Critically, overall response times did not differ between age groups, and the stronger alpha power reduction at trial ending for older participants was specific to the alpha frequency band (i.e., no motor-associated beta-band effect in a post-hoc analysis). As a consequence, the stronger alpha power decrease towards the end of a trial in older listeners was unlikely driven by an earlier response preparation. Instead, the reduced overall alpha power might reflect decreased maintenance of selective attention in older listeners (Gazzaley et al., 2005a). In line with this view, decreased lateralization of alpha power in older participants under high cognitive load has been interpreted as less efficient sustained inhibition of task-irrelevant neural processing (Sander et al., 2012).

In the behavioral results, we found that varying acoustic detail exerted a stronger relative impact on accuracy and response times in older adults. Thus, despite previous reports on reduced sensitivity to temporal fine structure variations in older adults (Grose and Mamo, 2010; Hopkins and Moore, 2011; Moore et al., 2012), older listeners in the current study relied relatively more on acoustic cues for their

performance. The strong dependence on acoustic cues in older listeners is in line with a stronger acoustics-driven decrease in alpha power after the presentation of the first digit in older participants (see also Sebastian et al., 2011). One attractive interpretation is that older listeners' attentional focus is more strongly affected by acoustic features of the external signal, potentially related to their difficulty in ignoring irrelevant auditory distractors (Chao and Knight, 1997; Tun et al., 2002; Passow et al., 2014). To our knowledge, there has been only one (behavioral) study that has shown a stronger dependence of speech recognition on spectral degradations at an older age, comparable to our observation (Schvartz et al., 2008). The present results thus demonstrate that age-related changes in listening behavior are reflected in neural alpha oscillations.

Notably, one rationale in the current study was to equalize audibility of materials (through individual control for frequency-specific audiometric thresholds) and the overall performance level (through individual adjustment of the digit-to-masker ratio) across participants to avoid propagated effects of hearing acuity on brain dynamics (Tremblay et al., 2003; Peelle et al., 2011). However, conventional auditory threshold measures do not capture all aspects of auditory processing acuity. For instance, age and noise-exposure might affect the neural encoding of supra-threshold sounds (Kujawa and Liberman, 2009; Ruggles et al., 2012; Furman et al., 2013) and could also contribute to observed age differences in listening behavior and alpha dynamics. Although sensory encoding is commonly impeded in older listeners, it is unclear whether this affects perception (Clinard et al., 2010; but see also Ruggles et al., 2012) and electrophysiological measures of cortical activity.

4.1.4.3 Alpha oscillations relate to subjective difficulty

We here extend previous findings of alpha oscillations as a neural marker of cognitive effort (e.g., Klimesch, 1999; Jensen et al., 2002) to one of the most common communication situations, that is, comprehending speech in multi-talker situations. Participants who showed weaker alpha power modulations with varying task difficulty reported higher difficulties of listening to speech in noise (Figure 4.5A). This is compatible with the view that higher neural variability accompanies enhanced behavioral performance (see Garrett et al., 2011; Erb and Obleser, 2013).

In addition to inter-individual differences in alpha power, we also found that *intra*-individual, trial-to-trial variations in alpha power affected post-trial confidence ratings: Lower alpha power during a trial predicted higher confidence of listeners in their own decision (Figure 4.5B). While correlations between alpha power and behavior have been found before (e.g., Klimesch et al., 1997; Haegens et al.,

2011a; Wilsch et al., 2014), the present changes in alpha power exerted an impact on subjective confidence ratings, a measure of so-called meta-cognition (Shea et al., 2014). Thus, fluctuations in alpha power not only reflect changes in the external stimulation, but they also constitute a change in brain state, which is independent of the stimulation yet can impact behavior (see Obleser and Weisz, 2012). The direction of the observed effect – lower alpha power for higher confidence ratings – supports the view that decreased alpha power reflects reduced task demands. These observations significantly extend the current understanding of alpha oscillations as a marker of subjective difficulty during effortful listening.

4.1.4.4 Conclusions

The current study shows that alpha oscillations support auditory processing in younger and older listeners in noisy environments in multiple ways. First, alpha oscillations are modulated by stimulation-related encoding demands induced by acoustic detail, but are also sensitive to the degree of stimulus predictiveness. Second, task performance and alpha modulation in older listeners are stronger affected by varying acoustic detail. This speaks to changes in attentional control at an older age. Lastly, alpha oscillatory dynamics explain inter- and intra-individual differences in introspective task demand. In sum, alpha dynamics are a promising neural marker to elucidate on individual and age-related difficulties in sensation, perception, as well as decision-making.

4.2 Study 2.2: Phase-locking of neural responses to attended and ignored speech

4.2.1 Introduction

In the previous chapter (4.1), we have shown that neural alpha oscillations (~ 10 Hz) signify listeners' attentional challenges in the auditory number comparison task. These alpha oscillations were however not phase-locked to the acoustic stimulation, meaning that there was no 10-Hz rhythm in the external acoustic signal that was picked up by these alpha oscillations. In the present chapter, we will present data from an alternative analysis of the same dataset. We will show in how far younger and older listeners' EEG responses phase-lock ("entrain") to the slow amplitude modulation (i.e., temporal envelope) of attended and unattended (i.e., ignored) speech signals.

The common EEG response to the presentation of an auditory (or visual) stimulus is the generation of an event-related potential (ERP) with particular waveform characteristics (for an example of such an ERP, see also Study 1). Across multiple stimulus presentations, ERP components occur at regular time points relative to stimulus onset (time-locking) and show a consistent waveform morphology (phase-locking). Thus, the ERP is a classic example for evoked (time- and phase-locked) activity in the EEG that can be contrasted with induced responses, which are not strictly time- or phase-locked (for a comprehensive illustration of evoked and induced activity, see Tallon-Baudry and Bertrand, 1999).

While the evoked response relative to a certain time point (e.g., stimulus onset) can be calculated by simple averaging in time over multiple stimulus presentations, the calculation of evoked neural activity in response to a continuous speech signal that unfolds in time over several seconds requires more sophisticated analysis techniques. Critically, it requires that the entire speech signal is related to the EEG response recorded during the presentation of that speech signal. One straightforward approach is to calculate the Pearson correlation of the EEG signal and the speech envelope. However, this correlation would not account for the fact that the EEG signal might phase-lock to the speech envelope with some time lag. Thus, the correlation for multiple time lags between speech envelope and EEG signal has to be calculated, using cross-correlation (for possible applications of cross-correlation in the neurosciences, see Salinas and Sejnowski, 2001). Here, we used cross-correlation to test whether the EEG signal differentially phase-locks to attended and ignored speech signals in younger as well as in healthy older listeners.

It has been long known that evoked responses are larger for attended compared to ignored auditory stimuli (e.g., Picton and Hillyard, 1974; Näätänen et al., 1981). Recently, it has been shown that also in

multi-talker situations, neural responses show a stronger phase-locking to the speech envelope of attended compared to ignored speech (e.g., Ding and Simon, 2012; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013; O'Sullivan et al., 2014). In the cross-correlation of EEG signal and speech envelope, prior studies consistently found a strong phase-locking with a time lag of approximately 100–200 ms (e.g., Hambrook and Tata, 2014; Kong et al., 2014). It has been suggested that the phase-locking of the EEG response to the speech envelope could serve to enhance neural excitability at time points critical for speech comprehension, such as the onset of syllables (Giraud and Poeppel, 2012). This hypothesis gains further support by studies that found opposite patterns of phase-locking for attended and ignored speech, suggesting that neural excitability might be enhanced for attended and simultaneously reduced for ignored speech signals (e.g., Horton et al., 2013).

Although EEG responses phase-lock to the slowly varying temporal envelope of speech signals, it has been shown that the preservation of fast fluctuations in the acoustic signal (i.e., fine structure) aids phase-locking to the envelope in background noise (Ding et al., 2013). This supports the often-postulated important role of fine structure for speech comprehension against fluctuating maskers such as speech (for a review on the role of fine structure, see Moore, 2008b). There is evidence that older listeners have a reduced sensitivity to fine structure (Grose and Mamo, 2010; Hopkins and Moore, 2011; Moore et al., 2012), which could partly explain older listeners' difficulties of speech comprehension in background noise (see Pichora-Fuller, 2003b). One possible underlying mechanism might be that older listeners' neural phase-locking to the speech envelope does not benefit from fine structure in speech.

In the present chapter, we asked whether EEG responses would show differential phase-locking to attended speech (i.e., spoken digits) and to the ignored speech masker in the auditory number comparison task. To extend findings of prior studies, we further tested to what extent phase-locking of the EEG signal to the speech envelope depends on the preservation of fine structure for younger and older listeners.

4.2.2 Methods

For all analyses described below, we used the same EEG dataset described extensively in Studies 1.1 and 2.1 (see sections 3.1 and 4.1). The data were recorded while 18 younger (20–30 years) and 20 older participants (60–70 years) performed an auditory number comparison task.

4.2.2.1 Phase-locking of EEG signals to speech envelopes

To extract acoustic envelopes of speech signals, we calculated the absolute of the Hilbert transform of spoken target digits and the irrelevant speech masker (Figure 4.6). Envelopes were lowpass-filtered at 25 Hz and down-sampled to the same sampling rate as the EEG signals (500 Hz). In order to emphasize amplitude changes in speech envelopes, the first derivative of envelopes was calculated. Finally, envelopes were half-wave rectified and normalized so that the summed amplitude across samples was equal to 1.

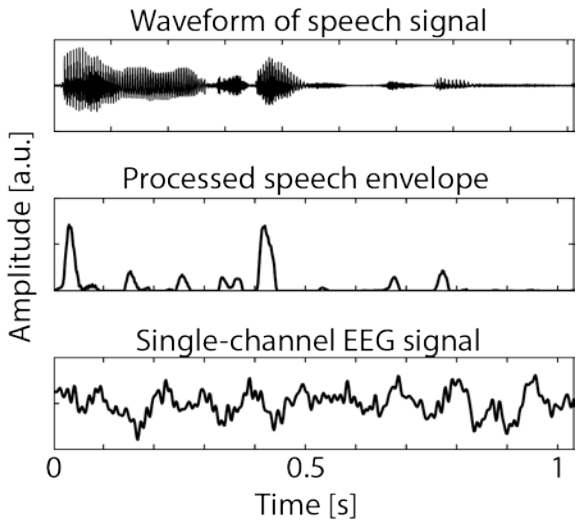

Figure 4.6. Example materials used for the calculation of neural phase-locking. Top panel shows the speech waveform of a spoken German digit ("61"). Middle panel shows the respective processed speech envelope after lowpass-filtering, down-sampling, calculation of the first derivative, and half-wave rectification. Note that amplitude changes in the speech waveform are emphasized in the processed speech envelope (e.g., syllable onsets). Bottom panel shows a single-channel EEG signal recorded simultaneously to the presentation of the speech signal. For the calculation of neural phase-locking, the cross-correlation of processed speech envelope and EEG signal was calculated.

For each trial in the auditory number comparison task, EEG signals were extracted during the presentation of target digits and speech masker. As a measure of neural phase-locking to speech envelopes, cross-correlations of EEG signals and speech envelopes at 28 scalp electrodes were calculated (using the *crosscorr* function in the Econometrics toolbox for Matlab, R2013b). The resulting correlation coefficients (r) as a function of time-lags were bound between –1 and 1. Since acoustic detail was manipulated over six levels in the auditory number comparison task (six temporal fine structure cut-off frequencies: 0, 0.1, 0.2, 0.4, 0.76, 1.45 kHz), cross-correlations were subsequently averaged over individual trials separately for each level of acoustic detail as well as attended speech signals (i.e., spoken digits) and ignored speech signals (i.e., speech maskers).

4.2.2.2 Attentional modulation

In order to test whether neural phase-locking to attended and ignored speech would differ, we calculated the difference between cross-correlations for attended and ignored speech (attended–ignored) for all participants and all levels of acoustic detail across eight fronto-central electrodes showing the strongest cross-correlations. To test whether acoustic detail would impact the attentional modulation, we calculated the linear coefficients, quantifying the change of attentional modulation over the six levels of acoustic detail (zero-centered predictor values: –2.5, –1.5, –0.5, 0.5, 1.5, 2.5) for the two peaks found in the cross-correlation (see below).

4.2.2.3 Auditory modelling

Since acoustic signals are decomposed into different frequency bands in the human cochlea, the question arises whether EEG signals phase-lock to speech envelopes in different frequencies to the same extent. To approach this question, we used the Auditory Modeling Toolbox (version 0.9.6; Sondergaard and Majdak, 2013) for Matlab. Acoustic signals were bandpass filtered into 29 logarithmically spaced frequency bands with centre frequencies between 60 and 6000 Hz (referred to as "cochlear filters" hereafter). The envelope in each of these 29 frequency bands was extracted using the Hilbert transform. Next, envelopes in all frequency bands were subjected to a modulation filterbank (Dau et al., 1997), which applied 12 logarithmically spaced bandpass filters with centre frequencies between 0 and 992 Hz to the data. Modulation filters capture the amplitude modulation (AM) of the speech envelope in different frequencies. There is evidence for the existence of modulation filters in the human auditory system (e.g., Langner and Schreiner, 1988; Jepsen et al., 2008; McDermott et al., 2013), although their neural implementation is not entirely clear (see Joris et al., 2004). Speech envelopes in the 29 (cochlear filters) × 12 (modulation filters) space were processed as described above (down-sampling, derivative, half-wave rectification) and the cross-correlation with the EEG signal was calculated. This analysis was purely exploratory and only performed for the EEG signal measured at a single frontal electrode (Fz), which showed the largest cross-correlation across experimental conditions.

4.2.3 Results

4.2.3.1 Differential phase-locking to attended and ignored speech

Figure 4.7A shows neural phase-locking to attended and ignored speech quantified by cross-correlation of speech envelope and the EEG signal. Cross-correlations for attended and ignored speech

exhibited an early positive deflection peaking at a time lag of 40 ms, and a later deflection peaking at 170 ms, which was negative for attended and positive for ignored speech. In the following, we will refer to these two peaks simply as *early* and *late* peaks, respectively. For statistical analysis, we averaged across time windows around the early (10–70 ms) and late peak (110–230 ms), as well as across eight fronto-central electrodes showing the strongest cross-correlations (see topographic maps in Figure 4.7A).

For the early peak, the average correlation coefficient was not significantly different between attended and ignored speech (t_{37} = 0.49; p = 0.628; r = 0.08). However, the average correlation coefficient for the late peak was significantly more negative for attended compared to ignored speech (t_{37} = 8.37; p < 0.001; r = 0.81), indicating differential neural phase-locking to attended compared to ignored speech. Average correlation coefficients for both peaks as well as for attended and ignored speech did not differ significantly between age groups (all p > 0.13; all r < 0.25).

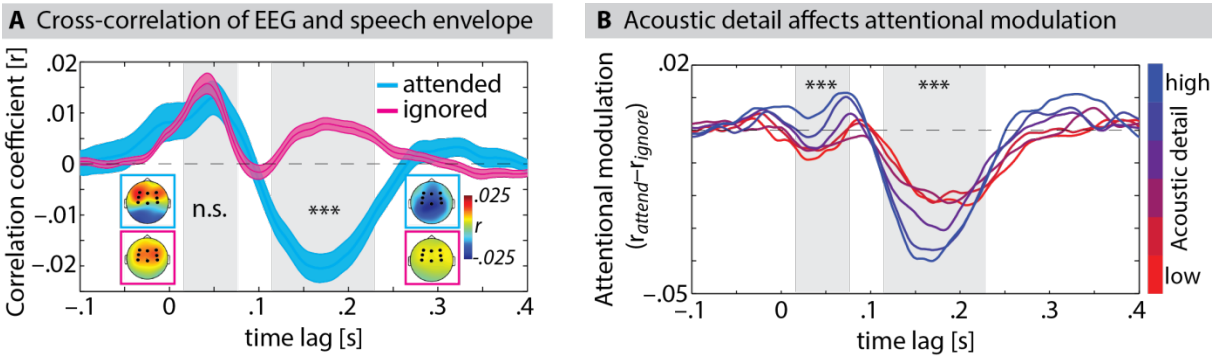

Figure 4.7. Cross-correlation of EEG signals and speech envelopes. (A) Cross-correlations of EEG signals with envelopes of attended speech (digits; cyan) and ignored speech (masker; magenta) in the auditory number comparison task. Cross-correlations were averaged across all (younger and older) participants and across eight fronto-central electrodes highlighted in the topographic maps. Time intervals around peaks in the cross-correlation, which were used for statistical analyses (attended vs ignored speech), are highlighted in light grey (early peak: 10–70 ms; late peak: 110–230 ms). Topographic maps show the average correlation coefficient for the two peaks (left: early peak; right: late peak) and for attended (cyan frame) and ignored speech (magenta frame). (B) Attentional modulation, calculated by subtracting the cross-correlation for ignored speech from attended speech. The plot shows the attentional modulation for six levels of acoustic detail (fine structure preservation below six frequency cut-offs: 0 (low), 0.1, 0.2, 0.4, 0.8, 1.5 kHz (high)). Asterisks indicate a significant linear change in attentional modulation over levels of acoustic detail. n.s., not significant; *** p < 0.001.

4.2.3.2 Acoustic detail affects attentional modulation of phase-locking

Figure 4.7B shows how attention modulated phase-locking to speech signals as a function of acoustic detail (temporal fine structure preservation). In detail, the attentional modulation was quantified by subtracting the cross-correlation for ignored from the cross-correlation for attended speech. The effect of acoustic detail was modelled by calculating participants' linear coefficients quantifying the change in average attentional modulation in the time intervals of the early and late peak (highlighted in grey in Figure 4.7) over six levels of fine structure preservation. Linear coefficient for the early peak were

significantly larger than zero (t_{37} = 5.10; p < 0.001; r = 0.64), indicating that the difference in phase-locking to attended and ignored speech was modulated by acoustic detail. For the late peak, linear coefficients were significantly smaller than zero (t_{37} = –6.20; p < 0.001; r = 0.71), indicating that the difference between phase-locking to attended and ignored speech increased with higher levels of acoustic detail. Linear coefficients for early and late peaks did not differ between age groups (both p > 0.5; both r < 0.11), suggesting a similar impact of fine structure on neural phase-locking across age groups.

4.2.3.3 Phase-locking to different frequencies in the speech signal

We performed an exploratory analysis on phase-locking of the EEG signal to the speech envelope as a function of cochlear filters frequencies and modulation filter frequencies. In essence, the same approach was used to calculate cross-correlations as shown above (Figure 4.7), but this time for speech signals divided into 29 cochlear frequency bands and speech envelopes divided into 12 modulation frequency bands (for details see Methods).

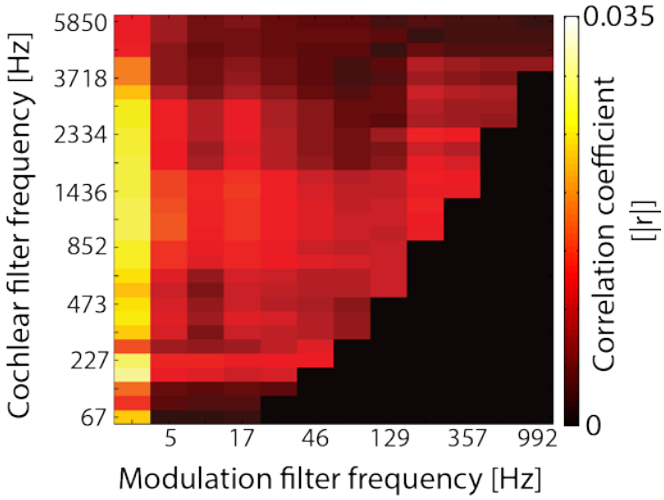

Figure 4.8. Neural phase-locking as a function of cochlear and modulation filter centre frequencies. The graph shows the average absolute cross-correlation across all participants, acoustic detail levels, speech materials (attended & ignored), and all time lags between 10 and 230 ms as a function of auditory filter frequencies (y-axis) and modulation filter frequencies (x-axis). The absolute correlation coefficients (|r|) were calculated to transform both positive (early peak) and negative correlation coefficients (late peak) into a composite positive value. Data in the lower right are missing since modulation filters were only applied to outputs of cochlear filters not including the centre frequency of modulation filters.

Since visual inspection of the resulting cross-correlations as a function of cochlear and modulation filter frequencies revealed very consistent effects across acoustic detail conditions as well as for attended and ignored speech, we collapsed across these dimensions. In detail, we calculated the absolute correlation coefficients across time lags to make sure that positive (early) and negative peaks (late) would not cancel out. Next, we averaged across time lags from the beginning of the early until

the end of the time period of the late peak (10–230 ms), as well as across acoustic detail conditions, attended and ignored speech, and across younger and older participants. Figure 4.8 shows the average absolute correlation coefficients as a function of cochlear and modulation filter frequencies. Correlation coefficients were highest for the lowest modulation filter frequencies (< 5 Hz) across all cochlear frequencies, indicating a strong neural phase-locking to slow envelope modulations of the speech signal. Correlation coefficients were also enhanced for cochlear filter frequencies around 1436 Hz. Above cochlear filter frequencies of 3718 Hz, phase-locking decreased.

4.2.4 Discussion

We investigated younger and older listeners' neural phase-locking to attended and ignored speech under different levels of acoustic detail (temporal fine structure). Results can be summarized as follows: (1) EEG responses showed distinct patterns of phase-locking to attended and ignored speech. (2) Across age groups, the effect of attention on neural phase-locking was increased with more acoustic detail in the speech material. (3) Neural phase-locking was strongest to slow (< 5 Hz) fluctuations of the speech envelope.

4.2.4.1 Neural phase-locking to attended and ignored speech

We found two dominant peaks in the cross-correlation of EEG signal and speech envelope (Figure 4.7A), an early one at a time-lag of 40 ms, and a late one at a time-lag of 170 ms. Latency and waveform characteristics of these peaks agree with prior studies analysing neural phase-locking by means of cross-correlation (Horton et al., 2013; Hambrook and Tata, 2014; Kong et al., 2014). Latency and polarity of these peaks resemble the common P1 and N1 event-related potential (ERP) components, respectively, occurring after the onset of acoustic events. These peaks in the cross-correlation are likely generated by the occurrence of an amplitude increase in the speech envelope, which is followed by a positive deflection in the EEG after 40 ms (P1) and by a subsequent negative deflection after 170 ms (N1). As Figure 4.6 shows, amplitude increases in the speech envelope are strongest at syllable onsets. Critically, the cross-correlation takes into account the entire speech envelope and the simultaneously recorded EEG signal. Thus, peaks in the cross-correlation not only reflect evoked responses to the onset of the speech signal (i.e., onset of first syllable) but rather a continuous phase-locking to the entire speech signal (i.e., all syllables). However, it must be mentioned that also different dependencies between speech envelope and EEG signal could generate the observed peaks in the cross-correlation. Future studies could use longer speech segments and calculate the

cross-correlation in consecutive time windows throughout the trial to uncover the temporal dependencies between EEG signal and speech envelope.

The late peak in the cross-correlation exhibited a strong negative correlation for attended speech, and a weaker positive correlation for ignored speech (Figure 4.7A). Differential neural phase-locking to attended and ignored speech has been shown previously using cross-correlations (e.g., Horton et al., 2013; Hambrook and Tata, 2014; Kong et al., 2014), as well as other methods incorporating spectro-temporal weighting of speech signals (e.g., Ding and Simon, 2012; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013; O'Sullivan et al., 2014). One possible interpretation of our findings rests upon the assumption that cortical oscillations measured in the EEG reflect fluctuations in neuronal excitability in auditory cortex (for a review on neuronal oscillations, see Buzsaki and Draguhn, 2004; Lakatos et al., 2005; Schroeder and Lakatos, 2009). Thus, phase-locking of EEG signals to speech envelopes might indicate that neural excitability fluctuations align with the speech envelope, possibly in such a way that high-amplitude parts of the signal (e.g., syllables) fall into high-excitability phases (Giraud and Poeppel, 2012). Critically, an opposite pattern of neural phase-locking to attended and ignored speech suggests that excitability is enhanced for attended but also lowered for ignored speech.

One shortcoming of the present analysis is the fact that the attended spoken digits were shorter (~1.13 s) compared to the ignored speech masker (~4.25 s) on each trial. Thus, more data samples were used to calculate the cross-correlation for the (ignored) masker compared to the (attended) digits. Possibly, this resulted in a better estimate of cross-correlations for ignored speech, which also explains the smaller standard error of the mean for the cross-correlation with ignored speech in Figure 4.7A. However, it is unlikely that this difference in the number of data samples was the reason for the strong difference in neural phase-locking to attended and ignored speech.

4.2.4.2 Fine structure aids differential phase-locking to attended and ignored speech

We found that the difference in neural phase-locking to attended and ignored speech (i.e., attentional modulation; Figure 4.7B) increased with more fine structure in speech materials. Prior work has shown that fine structure facilitates the perceptual segregation of acoustic signals (for review, see Moore, 2008b). Our finding is in line with one prior study which found more robust cortical entrainment to speech with more preserved fine structure (Ding et al., 2013). This is an interesting result since the envelope of speech, which was used for the calculation of neural phase-locking in the present study, is commonly thought to be largely independent of fine structure (for an opposing view,

see Shamma and Lorenzi, 2013). However, our results suggest that envelope and fine structure are both critical for neural phase-locking to speech.

Despite evidence for reduced sensitivity to temporal fine structure at an older age (Grose and Mamo, 2010; Hopkins and Moore, 2011; Moore et al., 2012), we found that preservation of fine structure increased the attentional modulation of phase-locking to the speech envelope similarly in younger and older adults. It might be that sensitivity to fine structure was not critically impaired in our sample of older listeners or that our individual adjustments of speech materials (see section 3.1.2.4) compensated for a reduced sensitivity to fine structure: It is currently under debate whether age-related cognitive decline or age-related hearing loss drives older listeners' decreased sensitivity to fine structure (e.g., Hopkins and Moore, 2011; Neher et al., 2012; Füllgrabe, 2013). Importantly, older listeners' preserved sensitivity to fine structure in the present study was also evidenced by their significant behavioural benefit from more fine structure in speech materials (see chapters 3.1 and 4.1). Since stimuli in the present study were carefully adapted to individual hearing acuity, speech materials for older listeners were amplified, especially in higher frequencies. Moreover, the digit-to-masker sound level ratio was higher for older participants (for details, see section 3.1.3.1). Possibly, older participants' benefit from fine structure would have been reduced without these stimulus adjustments. Thus, it might be that our individual adjustments compensated for a reduced sensitivity to fine structure in older listeners in the present study.

4.2.4.3 Neural phase-locking is strongest for slow envelope fluctuations

Our exploratory analysis of neural phase-locking across cochlear filter frequencies and modulation filter frequencies revealed that phase-locking is not equally strong across different frequencies of the speech signal (Figure 4.8). We observed the strongest neural phase-locking to slow (< 5 Hz) modulations in the speech envelope. The syllable rate of normal speech is between 3 and 6 Hz. Specifically, spoken digits in the present study contained four syllables and had an average duration of 1.125 sec, resulting in a syllable rate of 3.5 Hz. Thus, this finding suggests in line with prior research (Ahissar et al., 2001; Nourski et al., 2009; Giraud and Poeppel, 2012; Hertrich et al., 2012) that neural phase-locking to speech signals depends critically on the slow temporal envelope fluctuations at the syllable rate. Moreover, Figure 4.8 shows that neural phase-locking was strong for cochlear filter frequencies between 850 and 2300 Hz. This approximately matches the frequency region of strongest energy in human speech (~ 300-3000 Hz). Thus, unsurprisingly, neural phase-locking to speech depends particularly on those frequencies that dominate human speech sounds.

Based on these results, one might hypothesize that neural-phase locking would be most severely impaired by an acoustic masker with high energy between 850 and 2300 Hz as well as strong amplitude modulation of the envelope < 5 Hz. Future studies could directly test this hypothesis by measuring neural phase-locking to speech in the presence of an acoustic masker that varies systematically in its frequency range and envelope modulation rate.

4.2.4.4 Conclusions

Our results indicate robust neural phase-locking to the envelope of speech in younger as well as in healthy older listeners. Across age groups, temporal fine structure improves the separation of attended and ignored speech on a neural level. Thus, the present study shows that basic neural dynamics of auditory processing (i.e., phase-locking to the speech envelope) is preserved at an older age, given that acoustic conditions are carefully adjusted to individual hearing acuity.

5 Study 3: Dynamic lateralization of alpha power follows the speech rhythm and predicts successful attentional selection

This study investigates younger listeners' speech comprehension in a dichotic listening task. We asked in how far the hemispheric lateralization of the power of neural alpha oscillations reflects listeners' spatial attention to one of two concurrent speech streams.

5.1 Introduction

Processing relevant signals despite the presence of distraction (i.e., noise) requires selective attention. Spatial separation of signal and noise is a useful research paradigm to investigate the neural dynamics of attention to signals and suppression of noise. Spatial attention is reflected by an increase in the power of neural alpha oscillations (~10 Hz) in the ipsilateral hemisphere (same side as attended object) and a decrease in the contralateral hemisphere (opposite side as attended object). This alpha power lateralization has been found across sensory modalities for attention to visual (e.g., Thut et al., 2006; Bauer et al., 2012), somatosensory (e.g., Haegens et al., 2011a; van Ede et al., 2011), and auditory stimuli (e.g., Banerjee et al., 2011; Ahveninen et al., 2013). However, spatial attention paradigms typically involve only a single (brief) target stimulus and investigate neural activity only in anticipation of this target. Thus, the oscillatory dynamics of spatial attention to an ongoing signal that unfolds in time – such as human speech – are largely unknown.

According to the functional inhibition framework (Jensen and Mazaheri, 2010), alpha power regulates neural information flow through inhibition of task-irrelevant brain areas or processes. Prestimulus alpha power correlates negatively with the perception of near-threshold stimuli (e.g., Hanslmayr et al., 2007; van Dijk et al., 2008), suggesting that alpha power affects the degree to which stimuli become processed neurally. But does alpha power also regulate selective attention to ongoing stimuli in sensory-specific cortical areas? While alpha power modulation has been found in visual and somatosensory cortex regions (e.g., Haegens et al., 2011a; Spitzer et al., 2014), the existence and the functional significance of auditory alpha activity is unresolved (but see Lehtela et al., 1997; Müller and Weisz, 2012; Frey et al., 2014). In a recent perspective article, we proposed that alpha power modulation could serve as an attentional filter to enhance relevant and suppress irrelevant auditory input directly in auditory cortex regions (Strauß et al., 2014). Alpha power modulation in a parietal

attention network could serve the orienting of supramodal attention in space (Banerjee et al., 2011). In the present dichotic listening study, we used magnetoencephalography (MEG) to test whether alpha power lateralization indicates the direction (left vs right) of auditory selective attention to ongoing speech and we traced the underlying neural sources of this alpha power lateralization.

It is a common observation that the power of parietal alpha oscillations is enhanced under effortful listening conditions (e.g., Obleser et al., 2012; Wilsch et al., 2014; Wöstmann et al., 2015), which might reflect increased attention to the auditory modality (Adrian, 1944; Mazaheri et al., 2014). Furthermore, there is initial evidence that alpha power lateralization during spatial attention indicates the attentional selection of ongoing speech (Kerlin et al., 2010). However, it is unknown whether alpha power lateralization aligns with the inherent temporal regularity of the speech signal to regulate selective attention: The strongest attentional selection should co-occur with task-relevant speech items (such as individual words), whereas selective attention could be reduced in uninformative time intervals between speech items. This would result in an oscillation between states of high and low selective attention. Although there is evidence that attention modulates the alignment of certain neural responses with speech signals (e.g., Ding and Simon, 2012; Hertrich et al., 2012; O'Sullivan et al., 2014; see also Study 2.2), it has not been investigated whether the lateralization of alpha power aligns with the speech signal to support selective attention. Here we will demonstrate that alpha power lateralization during spatial attention to one of two speech streams temporally aligns with the word rate and that this alignment predicts the success of auditory stream selection.

5.2 Methods

5.2.1 Participants

Nineteen young (mean age = 27.47; age range = 23–34 years; 10 females) right-handed German native speakers participated in this study. Data of one additional participant were recorded but excluded from all analyses due to technical problems during MEG recording. Participants were financially compensated for participation. Procedures were approved by the local ethics committee of the University of Leipzig Medical faculty.

5.2.2 Auditory materials

We used the same recordings of German spoken digits between 21 and 99 (excluding integer multiples of ten) as in Studies 1.1 and 2.1 (Wöstmann et al., 2015; Wöstmann et al., in press). Digits were spoken

by a trained female speaker and recorded at a sampling rate of 44.1 kHz. Each digit contained four syllables and digits were on average 1.125 s long. Intensity of digits was equalized to –30 dB FS (full scale). Since the temporal alignment of digits presented simultaneously to the two ears is essential in dichotic listening studies, we aligned the perceptual onsets of digits which are different from their acoustic onsets (Morton et al., 1976). To this end, we extracted the envelope of spoken digits (using the Hilbert transform) and applied a lowpass filter at 15 Hz. For each digit, we determined the time point where the envelope increase of the first syllable reached 50 % of the syllable's maximum amplitude. This time point was considered the perceptual onset of the respective digit and will be referred to as *digit onset* hereafter.

For the background noise, we generated random noise at an intensity of –40 dB FS. Thus, the signal-to-noise ratio (SNR) between spoken digits and background noise was +10 dB. Different instances of noise were generated for each experimental trial. For the spatial cue at the beginning of each trial, we generated a 1000-Hz pure tone of 500 ms duration at an intensity of –30 dB FS (equal intensity as spoken digits).

5.2.3 Procedure of the dichotic listening task

Each trial started with the binaural presentation of background noise and a simultaneous spatial cue (i.e., pure tone) either to the left side (i.e., left ear) or right side (i.e., right ear). Linear onset ramps of 50 ms duration were applied to background noise and spatial cue. 1.8 s after cue offset, two streams of four spoken digits were presented simultaneously to the left and right side (Figure 5.1). Simultaneously presented digits were always distinct in their first (i.e., tens) and second (i.e., ones) position (e.g., combinations of "35" and "37" or "81" and "21" were avoided). The onset-to-onset time delay between subsequent digits on each side was 1.49 s, resulting in a digit presentation rate of 0.67 Hz on both sides (cp. Gomez-Ramirez et al., 2011; Lakatos et al., 2013a). During acoustic stimulation, participants fixated a centrally presented cross. Participants' task was to attend and to retain the digits presented on the side where the spatial cue had appeared and to ignore digits on the other side.

A response screen appeared 0.8–1.2 s (average 1 s) after the offset of the last digit. The response screen contained 12 digits, four from the attended side, four from the ignored side, and four random digits not presented on either side. To prevent participants' motor preparation during the trial, digits were presented either in ascending or descending order (randomized). Participants used an MEG-compatible trackball mouse (Logitech Marble Mouse) to select four digits which they thought had

been presented on the attended side. Individual digits disappeared from the screen after they had been selected. After the selection of the fourth digit, participants saw a pause screen. The next trial started self-paced, 1 s (randomly jittered between 0.8 and 1.2 s) after an additional mouse click. Auditory materials were presented via plastic ear tubes at an average intensity of ~70 dB SPL. Visual stimuli were shown on a back projection screen.

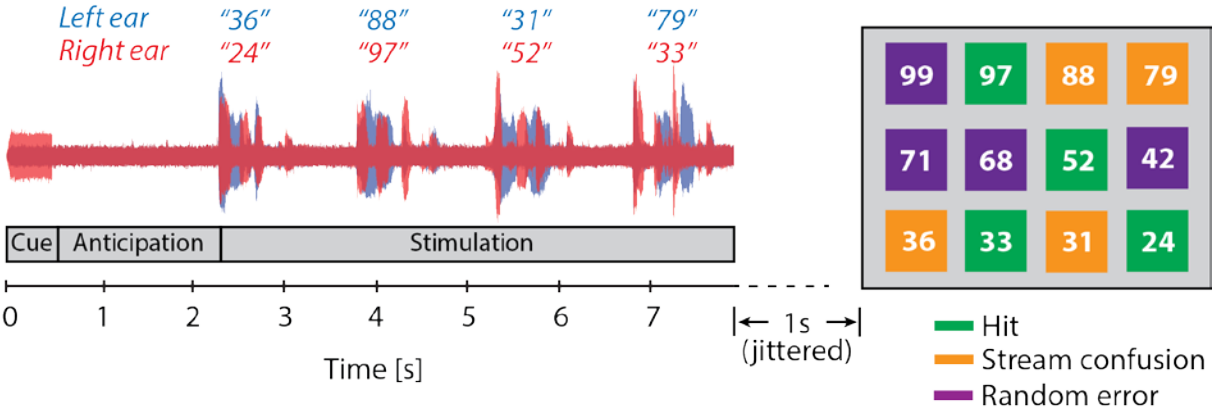

Figure 5.1. Trial design of the dichotic listening task. During the cue period (0–0.5 s), a pure tone was presented either on the left or right side (i.e., left or right ear) to indicate which side participants should attend (right side in this example). After an anticipation period (0.5–2.3 s), two streams of four spoken digits each were presented simultaneously on both sides during the stimulation period (2.3–7.9 s). All materials were presented in broadband background noise (+10 dB SNR; same noise presented on both sides). After acoustic stimulation, participants had the task to select digits from the attended side from an array of 12 digits (grey box on the right). Each response (selected digit) could either be a hit (digit appeared on attended side; green), a stream confusion (digit appeared on ignored side; orange) or a random error (digit did not appear on either side; purple). Coloured boxes indicate response types for the example trial depicted here. In the experiment, digits were shown in black font in white boxes.

Each participant performed 150 trials. The experiment was divided in 5 blocks of 30 trials each. Within each block, the spatial cue appeared on the left side in half of the trials and on the right side in the other half of trials. Trial order within each block was completely randomized. The entire experiment lasted approximately one hour.

5.2.4 Behavioral data recording and analysis

Due to technical reasons, no behavioral responses were recorded on 1.2 % (SD = 2.7 %) of trials across 19 participants. These trials were removed from all further behavioral and MEG data analyses. On remaining trials, we recorded four responses (four mouse clicks to indicate which digits appeared on the attended side). Responses were categorized in the following manner: A response was considered a *hit* if the selected digit had appeared on the attended side, a stream *confusion* if the digit had appeared on the ignored side, and a *random error* if the digit had appeared on neither side (Figure 5.1).

For statistical analyses, we calculated the proportion of different response types (hit, stream confusion, random error) through division by the total number of responses for each participant. To

ensure normal distribution of the data, we transformed the proportion data to rationalized arcsine units (rau), ranging between –.23 and 1.23 (Studebaker, 1985). Rau-transformed proportions of different response types were submitted to a repeated-measures ANOVA with the within-subject factor response type (hit, stream confusion, random error). Post-hoc paired t-tests were used for pairwise comparison between response types.

5.2.5 MEG data recording and analysis

Participants were seated in a magnetically shielded room (Vaccumschmelze, Hanau, Germany). A 306-sensor Neuromag Vectorview MEG (Elekta, Helsinki, Finland) was used to measure magnetic fields at 102 locations from 204 orthogonal planar gradiometers and 102 magnetometers. Additionally, we recorded the electroencephalogram (EEG) from 64 scalp electrodes (Ag/Ag-Cl). EEG data were not further analyzed in this study. Each participant's head position was monitored with five head position indicator (HPI) coils. MEG signals were recorded at a sampling rate of 1000 Hz with a DC–330 Hz bandwidth. Offline, the signal space separation (SSS) method (Taulu et al., 2004) was applied to suppress external disturbances (i.e., noise) in the data, to interpolate bad sensors, and to transform individual data to a common sensor space allowing for statistical comparison across participants.

For all subsequent MEG data analyses, we used the Fieldtrip toolbox (Oostenveld et al., 2011) for Matlab (R2013b) and customized Matlab scripts. For all analyses, we used only data recorded from gradiometer sensors. Continuous data were highpass filtered at 0.3 Hz using a causal finite impulse response (FIR) filter with time correction and lowpass filtered at 180 Hz using an acausal (bidirectional) FIR filter. Epochs from –2 to 10 s around cue onset were extracted from the continuous data. Data from five experimental blocks were appended and down-sampled to 500 Hz. Epochs were rejected when the signal at any sensor exceeded 800 pT/m. An independent component analysis (ICA) was performed and components corresponding to eye blinks, saccadic eye movements, muscle activity, heartbeats, drifts, and jumps were identified and rejected by inspection of components' topographies, time courses, and frequency spectra.

Time-frequency representations of single trials were estimated by convolving the single-trial time series with a family of Morlet wavelets between 1 and 20 Hz (in steps of 0.5 Hz; width: 7 cycles) and from –1.7 to 9 s (in steps 0.05 s). Single-trial power was obtained by squaring the magnitude of the estimated complex wavelet transform coefficients. Data from 204 gradiometer sensors (102 pairs of gradiometer sensors) were combined by summation of power estimates from the two sensors at the

same location. Thus, 102 (combined) gradiometer sensors were used for all subsequent sensor-space analyses.

5.2.6 Alpha power lateralization

To quantify the lateralization of alpha power as a function of participants' attention to digits on the left or right side, we contrasted participants' alpha power at 102 sensors, frequencies from 8 to 12 Hz, and all time points from 0 s (cue onset) until 7.9 s (last digit offset) in the following way: (attention left – attention right)/(attention left + attention right)(Figure 5.3). The resulting values (bound between –1 and 1) will be referred to as *alpha power lateralization* in the following. Values > 0 indicate higher alpha power when attention was directed to the left compared to the right side, and vice versa for values < 0. For statistical analyses, we compared the average alpha power lateralization between all sensors on the left and right hemisphere in three time periods (cue: 0–0.5 s; anticipation: 0.5–2.3 s; stimulation: 2.3–7.9 s) using paired t-tests. Since the topographical distribution of the alpha power lateralization differed between individuals, we selected 20 (combined gradiometer) sensors on the left hemisphere showing the most positive alpha lateralization, and 20 (combined gradiometer) sensors on the right hemisphere showing the most negative alpha power lateralization during the time of the entire trial (0–7.9 s) for each participant (Figure 5.3B). These individually selected sensors were used for the computation of the alpha lateralization index (see below).

5.2.7 Alpha lateralization index

In order to contrast the alpha power lateralization between correct (0 errors) and incorrect trials (1–4 errors), we calculated an *alpha lateralization index*. To this end, we contrasted alpha power at individually selected ipsilateral sensors (IPSI; same hemisphere as attended side) with alpha power at individually selected contralateral sensors (CONTRA; opposite hemisphere of attended side) in the following way: (ISPI – CONTRA)/(IPSI + CONTRA), separately for correct and incorrect trials (Figure 5.5A). Note that due to the selection of ipsi- and contralateral channels, we could aggregate trials across both conditions (attend left/right). For statistical analysis, we compared the average alpha lateralization index in three time periods (cue, anticipation, stimulation) between correct and incorrect trials using paired t-tests.

Since the alpha lateralization index exhibited characteristic fluctuations during the presentation of spoken digits (stimulation period: 2.3–7.9 s; Figure 5.5A), we quantified these fluctuations by fitting a cosine function to the alpha lateralization index. The cosine function [$g(t) = A \times \cos(2\pi \times f \times t + \phi)$]

had the fixed parameters time (t; 2.3–7.9 s) and frequency (f; 0–2 Hz in steps of 0.01 Hz) and the free parameters cosine amplitude (A) and cosine phase (ϕ). For each participant and frequencies 0–2 Hz, we fitted two cosine functions; one to the average lateralization index across correct trials and one to incorrect trials (using the lsqcurvefit function in the optimization toolbox for Matlab). Prior to fitting, the linear trend was removed from each participant's lateralization index. For statistical analysis, we were mainly interested in the cosine amplitude parameter (A), which quantifies the strength of rhythmic fluctuations of the alpha lateralization index. We compared the average cosine amplitude (A) for correct and incorrect trials in the frequency range of the 0.67-Hz digit presentation rate (averaged across 0.5–0.7 Hz) with a paired t-test (Figure 5.5C). To test for a phase effect, we computed the average circular distance between correct and incorrect trials (across 0.5–0.7 Hz), which was tested for non-uniformity (using a Rayleigh test implemented in the circular statistics toolbox for Matlab).

The number of trials was not balanced between correct (0 errors) and incorrect trials (1–4 errors; Figure 5.2C). In theory, this inequality in the number of trials might have affected our estimates of cosine amplitude (see above). However, the comparison of 95 % confidence intervals for the estimates of cosine amplitude at 0.67 Hz revealed no significant difference between correct and incorrect trials (t_{18} = 1.06; p = 0.303; r = 0.24). This suggested that our estimates of cosine amplitude were not significantly affected by the number of correct and incorrect trials. Nevertheless, we conducted an additional analysis in which we equalized the number of correct and incorrect trials. In detail, for a participant with fewer correct than incorrect trials, we selected (randomly without replacement) as many incorrect trials as there were correct trials, and vice versa for a participant with fewer incorrect than correct trials. Next, we calculated the lateralization index for these stratified samples of correct and incorrect trials and estimated the cosine amplitude of the alpha lateralization index (in the same way as described above). This procedure was repeated 1000 times for each participant. Finally, the mean cosine amplitude across 1000 repetitions and across frequencies 0.5–0.7 Hz was compared between correct and incorrect trials using a paired t-test.

Lastly, we tested in how far the amplitude of 0.67-Hz fluctuations of the lateralization index in incorrect trials predicts performance. Therefore, we calculated the correlation of cosine amplitude in incorrect trials (averaged across 0.5–0.7 Hz) and the average number of errors in incorrect trials across participants (Figure 5.5D). Since both variables entered in the correlation deviated significantly from the normal distribution (Lilliefors test; both $p < 0.05$) a nonparametric Spearman correlation was used.

5.2.8 Source analysis

Individual T1-weighthed MRI images were used to construct cortical surfaces and inner skull surfaces (using Freesurfer and MNE software). Individual cortical surfaces were used as source model and reduced to 10,242 vertices for each hemisphere. The inner skull surface served as volume conductor. The MR and the MEG coordinate systems were co-registered using 5 HPI coils and 64 digitized points on the head surface. Leadfields were calculated (i.e., the forward solution) on the basis of the cortical surfaces and inner skull (Nolte, 2003).

For the source localization of alpha power lateralization, we applied the Dynamic Imaging of Coherent Sources (DICS) beamformer approach (Gross et al., 2001) implemented in the Fieldtrip toolbox for Matlab. In detail, a spatially adaptive filter was used to estimate alpha oscillatory activity at the 10,242 source locations. We calculated Fourier spectra centered at 10 Hz with ±2 Hz spectral smoothing (8–12 Hz) separately for attention left and attention right trials and for three time periods (cue: 0–0.5 s, anticipation: 0.5–2.3 s, stimulation: 2.3–7.9 s). A complex common spatial filter was calculated on the basis of the Fourier spectra of trials in all conditions (attention left/right) from trial onset (0 s) to trial offset (7.9 s). The common filter was then used for source projection of attention left and attention right conditions separately (in the three time periods). Alpha power lateralization was calculated at each source location: (attention left – attention right)/(attention left + attention right). The resulting maps were spatially smoothed across the surface using an approximation to a 6 mm FWHM Gaussian kernel (Han et al., 2006), and the individual source estimates were morphed onto the cortical surface of one participant (Fischl et al., 1999a; Fischl et al., 1999b). Finally, for each source location, alpha power lateralization was tested against zero using a one-sample t-test. For visualization, the resulting t-values were transformed to z-values and overlaid on the partially inflated brain surface (Figure 5.4).

Furthermore, we estimated the sources of auditory activation. To this end, we determined participants' evoked oscillatory activity. In detail, we calculated the Fourier spectra of the average time-domain data across all trials (i.e., event-related fields) at the onset of the first digit (2.3–2.8 s) centered at 4 Hz with ±2 Hz spectral smoothing (2–6 Hz). A common spatial filter was calculated on the basis of the Fourier spectra from 1–7.9 s relative to cue onset. Note that the first second of trials was not used since the cue (0–0.5 s) appeared either on the left or right side on each trial, which could bias source estimation. To estimate the distribution of auditory activation, we calculated the neural activity index

(NAI) through division of the activation by the noise power estimate at each source location (Van Veen et al., 1997).

5.2.9 Granger causal influence

We asked whether the most central neural measure in the present study – the alpha lateralization index (Figure 5.5) – could be predicted from two other neural measures (in sensor-space), namely overall alpha power and inter trial phase coherence (ITPC). To this end, we averaged alpha power and the alpha lateralization index across all trials (attention left/right; correct/incorrect) within each participant. ITPC was computed through division of the complex wavelet coefficients obtained in the spectral analysis of MEG data (see above) by their absolute values and subsequent averaging across trials. ITPC is bound between 0 and 1 with higher values indicating a stronger phase-consistency of neural oscillations across individual trials. For the present analysis, we averaged each participant's ITPC across frequencies from 2–8 Hz, as well as across 6 left central combined gradiometer sensors and 6 right central combined gradiometer sensors that showed the largest ITPC at the onsets of auditory events (i.e., spoken digits; Figure 5.6A).

To investigate the directional dependence between these three neural measures (ITPC, alpha power, alpha lateralization index), we computed frequency-domain Granger causality (using the ft_connectivityanalysis function in the Fieldtrip toolbox). Linear trends were removed from ITPC, alpha power, and alpha lateralization index for each participant to ensure stationarity (Seth et al., 2015). We fitted an autoregressive model to each participant's data comprising the three neural measures during stimulation (2.3–7.9 s; using the bsmart toolbox for Matlab). Essentially, an autoregressive model explains how time domain data linearly depends on its own as well as other signals' past. The model order was 10, meaning that 10 past data samples were included in the model, corresponding to 500 ms given our temporal resolution of 50 ms. Granger spectra were computed with a frequency resolution of 0.1 Hz from 0–4 Hz (Figure 5.6B). Simply speaking, high values of granger causal influence from data A on B indicate that past values of A predict present values of B more than past values of B alone. In such a scenario, A is said to "granger cause" B. For statistical analysis, we compared granger causal influences in both directions (A→B; B→A) for each pair of the three neural measures (ITPC, alpha power, alpha lateralization index) using Wilcoxon signed-rank tests (nonparametric tests were used as the data deviated significantly from the normal distribution; Lilliefors test, all $p < 0.05$).

5.2.10 Effect sizes

To estimate effect sizes for F-statistics (ANOVAs), we calculated the partial eta-squared (η^2_P). Partial eta-squared values of 0.01, 0.06 and 0.14 indicate small, medium, and large effects, respectively (Cohen, 1969). For t-statistics (t-tests) and z-statistics (Wilcoxon signed-rank tests), we calculated the effect size measure r, which is bound between 0 and 1 (Rosenthal, 1994). For circular statistics (Rayleigh test), we computed the resultant vector length (r; bound between 0 and 1).

5.3 Results

5.3.1 Performance in the dichotic listening task

Figure 5.2A shows rau-transformed proportions of three different response types (hit, stream confusion, random error) in the dichotic listening task. A repeated-measures ANOVA revealed a significant main effect response type ($F(2, 18.6) = 286.54$; $p < 0.001$; $\eta^2_P = 0.94$; Greenhouse-Geisser correction applied due to violation of sphericity; Mauchly test: $p < 0.001$). Post-hoc tests revealed that the proportion of hits was higher than the proportion of stream confusions ($t_{18} = 15.87$; $p < 0.001$; $r = 0.97$) and random errors ($t_{18} = 18.26$; $p < 0.001$; $r = 0.97$). Critically, also the proportion of stream confusions was higher than that the proportion of random errors ($t_{18} = 7.05$; $p < 0.001$; $r = 0.86$; Figure 5.2B), indicating that participants erroneously reported digits on the ignored side more often than digits that did not appear on either side.

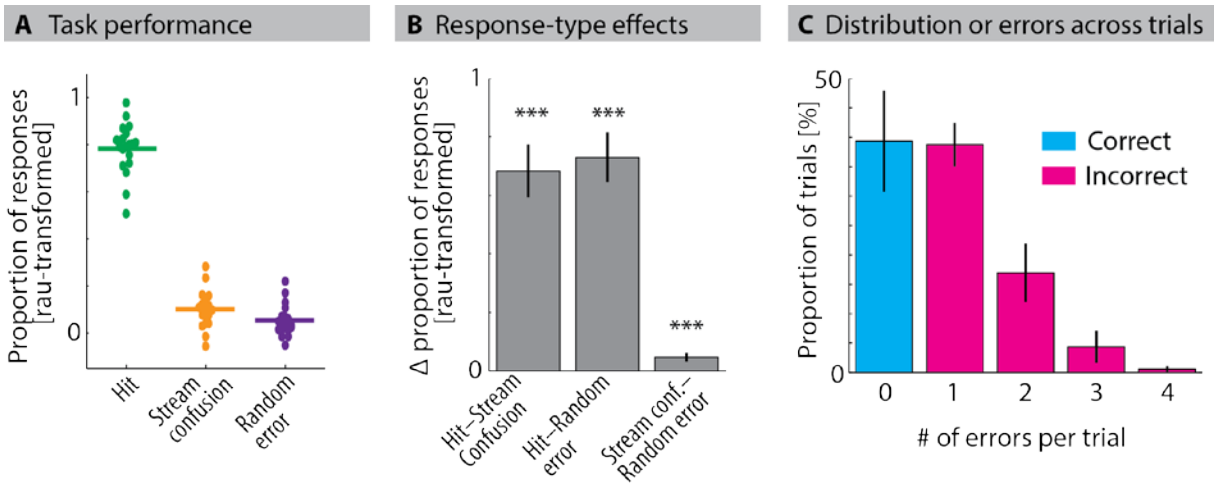

Figure 5.2. Behavioural performance in the dichotic listening task. (A) Proportion of response types (hit, stream confusion, random error) were transformed to rationalized arcsine units (rau). Dots show data of individual participants, horizontal lines show the average across participants. (B) Bars indicate the average difference between the proportions of all pairs of response types. (C) Bars indicate the average proportion of trials as a function of the number of possible errors on each trial (0–4). All error bars indicate 95 % confidence intervals. *** $p < 0.001$.

Figure 5.2C shows the proportion of trials as a function of the possible number of errors (0–4) on each trial. For further analyses of MEG data (see below), we contrasted correct trials (0 errors) with incorrect trials (1–4 errors).

5.3.2 Alpha power lateralization

Figure 5.3A shows oscillatory power averaged across 102 combined gradiometer channels, experimental conditions (attention left/right), and 19 participants. Power in the alpha frequency band (8–12 Hz) was prominently enhanced, and decreased gradually toward the end of the dichotic listening task. Alpha power lateralization (quantified as (attention left – attention right)/(attention left + attention right)) revealed a distinction of alpha power over the left and right hemispheres (Figure 5.3B): In attention left trials, alpha power was relatively higher over the left (ipsilateral) and lower over the right (contralateral) hemisphere compared to attention right trials.

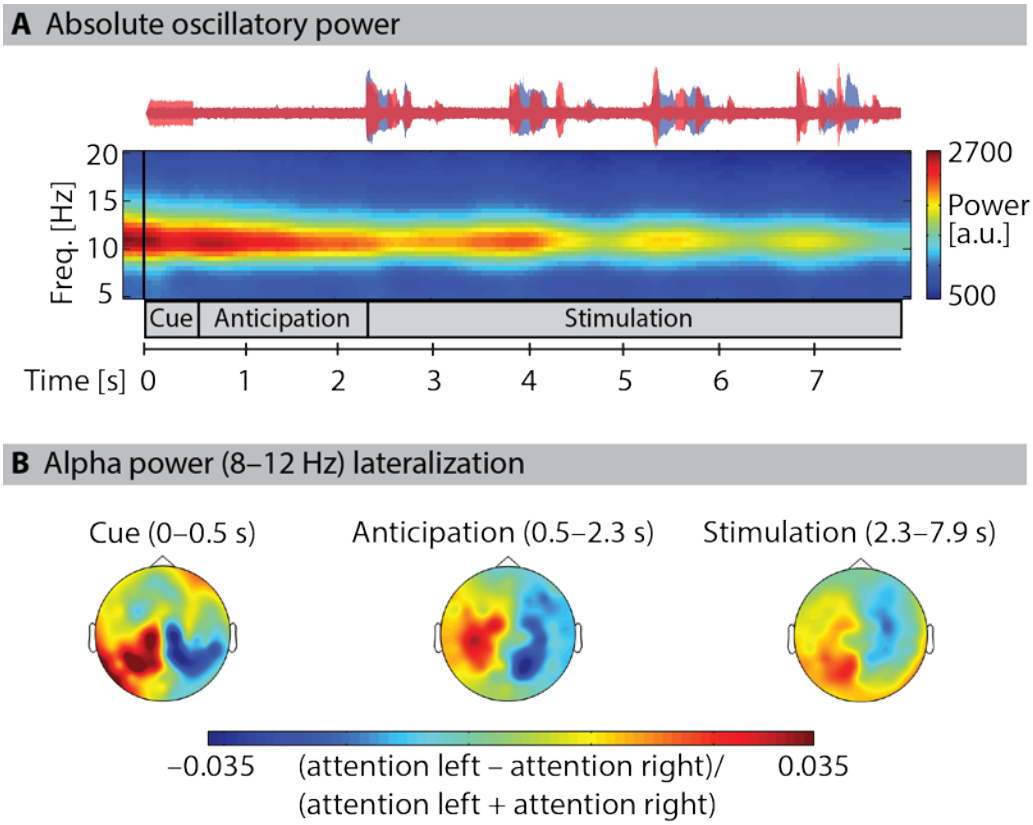

Figure 5.3. Alpha power lateralization in the dichotic listening task. (A) Grand average oscillatory power across 102 combined MEG gradiometer channels, both experimental conditions (attention left/right), and 19 participants. Note particularly high power in the alpha frequency band (8–12 Hz). (B) Alpha power lateralization was quantified by contrasting the two experimental conditions in the following way: (attention left – attention right)/(attention left + attention right). Resulting values are bound between –1 and 1. Topographic maps show relatively higher alpha power on the left hemisphere and lower alpha power on the right hemisphere for attention left compared to attention right trials in cue, anticipation, and stimulation period.

Alpha power lateralization showed significantly larger values on the left hemisphere than on the right during the cue period ($t_{18} = 2.5$; $p = 0.023$; $r = 0.51$), anticipation period ($t_{18} = 3.56$; $p = 0.002$; $r =$

0.64), and critically also during dichotic listening to spoken digits (i.e., stimulation period; $t_{18} = 2.55$; $p = 0.02$; $r = 0.52$). Thus, lateralized alpha power significantly differentiated between attention left and attention right trials.

5.3.3 Neural sources of power alpha lateralization

Figure 5.4 shows results of the beamformer source localization of oscillatory activity. As expected, auditory activation (2–6 Hz evoked oscillatory activity at the onset of the first digit) was localized to auditory cortex areas (Figure 5.4, left column).

The three columns on the right in Figure 5.4 show the source localization of alpha power lateralization, quantified as (attention left – attention right)/(attention left + attention right). Similarly to the estimation of alpha lateralization in sensor space, the neural sources of alpha lateralization were analysed separately for the three time periods cue (0–0.5 s), anticipation (0.5–2.3 s), and stimulation (2.3–7.9 s). For visualization, we show only those source estimates of alpha power lateralization that differed significantly from zero ($|z| > 1.96$). Figure 5.4 shows that the localization of alpha lateralization generally matched the sensor space data (cp. Figure 5.3B), with alpha lateralization values > 0 (attention left > attention right; red and yellow) on the left hemisphere and values < 0 (attention left < attention right; blue and cyan) on the right hemisphere. In the cue period, alpha lateralization was mainly localized to parietal cortex in the left hemisphere and to inferior parietal and parieto-occipital cortex in the right hemisphere (Figure 5.4, second column from left). In the subsequent anticipation period, inferior frontal cortex regions on the left and superior temporal as well as middle frontal cortex regions on the right hemisphere contributed additionally to the alpha power lateralization (Figure 5.4, third column from left).

Most critical for the present study were sources of alpha power lateralization in the stimulation period, where participants attended spoken digits on one side while they ignored digits on the other side. In the stimulation period, insula and parieto-occipital cortex on the left hemisphere exhibited the strongest increase in alpha power in attention left compared to attention right trials. On the right hemisphere, the increase in alpha power in attention right compared to attention left trials was mainly localized to auditory cortex (superior temporal cortex) and inferior parietal cortex. Importantly, the localization of alpha lateralization in the stimulation period overlapped with the localization of the peak auditory activation in the left and particularly in the right hemisphere (Figure 5.4, white outlines).

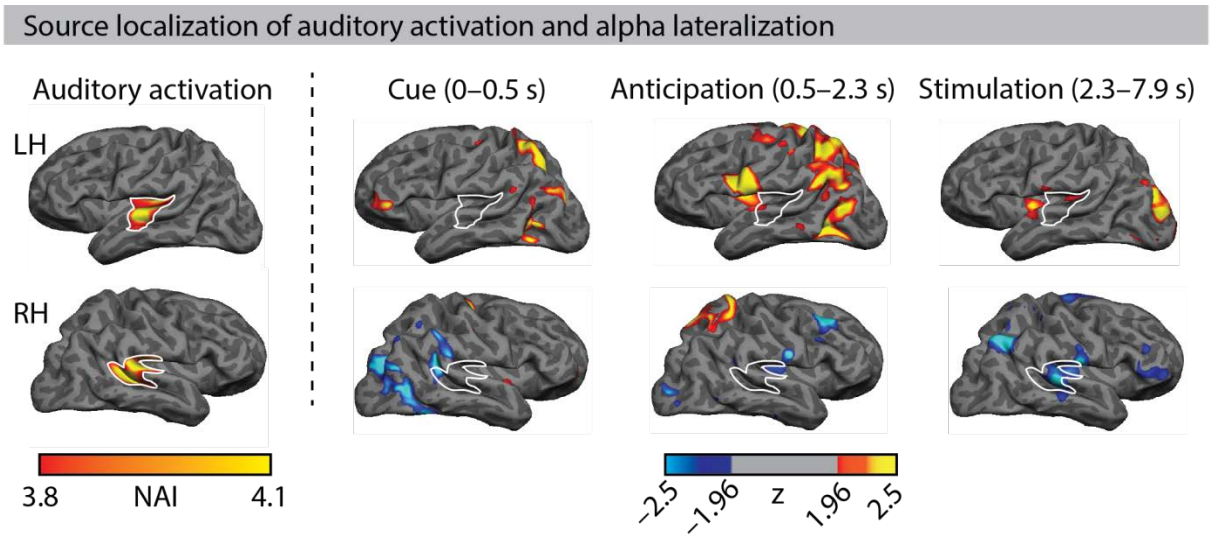

Figure 5.4. Source localization of oscillatory activity. (Left column) Neural activity index (NAI) of 2–6 Hz evoked oscillatory activity at the onset of the first digit (2.3–2.8 s). Three columns on the right show source estimates of the alpha power (8–12 Hz) lateralization – quantified by (attention left – attention right)/(attention left + attention right) – in cue, anticipation, and stimulation period thresholded at a z-value of ±1.96. White outlines indicate area of peak auditory activation.

5.3.4 Alpha lateralization aligns with the external stimulation

To quantify the lateralization of alpha power (in sensor-space) independent of the contrast between attention left and attention right trials, we calculated the alpha lateralization index. This index contrasts alpha power at sensors on the same hemisphere as the side of attention (ipsilateral) with sensors on the opposite side (contralateral). A positive index shows higher alpha power at ipsilateral compared with contralateral sensors. The average alpha lateralization index for correct trials (0 errors) and incorrect trials (1–4 errors) is shown in Figure 5.5A. The average lateralization index did not differ significantly between correct and incorrect trials in the cue period (0–0.5 s; $t_{18} = 0.27$; $p = 0.79$; $r = 0.06$), but was significantly enhanced for correct compared to incorrect trials in the anticipation period (0.5–2.3 s; $t_{18} = 2.84$; $p = 0.011$; $r = 0.56$). In the stimulation period, the increase of the alpha lateralization index for correct compared to incorrect trials approached statistical significance ($t_{18} = 1.84$; $p = 0.083$; $r = 0.4$).

Critically, the lateralization index was not constant over time but exhibited regular fluctuations, especially during the acoustic stimulation (Figure 5.5A; 2.3–7.9 s). The amplitude of these fluctuations at frequencies from 0 to 2 Hz was estimated by fitting cosine functions separately to the lateralization index in correct and incorrect trials for each participant. Figure 5.5C shows the average estimated cosine amplitude as a function of cosine frequency. Cosine amplitude at frequencies around the 0.67-Hz digit presentation rate (averaged across 0.5–0.7 Hz) was significantly higher in correct compared to incorrect trials ($t_{18} = 3.38$; $p = 0.003$; $r = 0.62$; Figure 5.5C inset). Note that this difference was not

abolished when we controlled for the unbalanced number of correct and incorrect trials through stratified sampling (see Methods; $t_{18} = 1.97$; $p = 0.064$; $r = 0.42$).

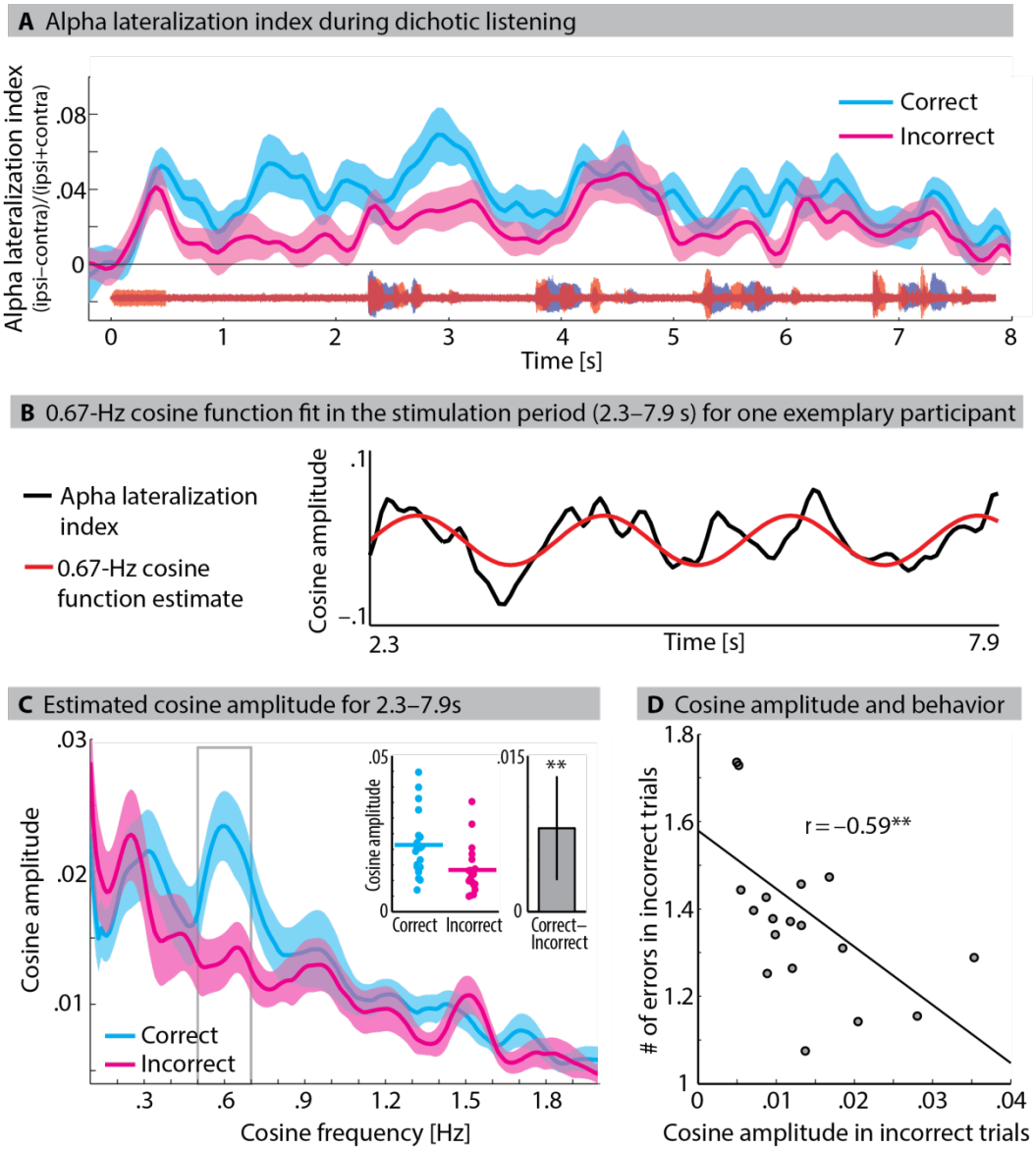

Figure 5.5. Fluctuating alpha lateralization index in the dichotic listening task. (A) Alpha lateralization index contrasts ipsi- and contralateral sensors (relative to the side of attention; left/right) in the following way: (ipsilateral – contralateral)/(ipsilateral + contralateral). Indices > 0 indicate larger relative alpha power at ipsi- compared to contralateral sensors. The graph shows the lateralization index separately for correct trials (0 errors; cyan) and incorrect trials (1–4 errors; magenta). Shaded areas indicate ±1 SEM. (B) Alpha lateralization index (black) and cosine function fit for a frequency of 0.67 Hz (red) for one exemplary participant in the stimulation period (2.3–7.9 s). (C) Fluctuations of the alpha lateralization index during stimulation (2.3–7.9 s) were quantified by fitting cosine functions at frequencies 0–2 Hz to each participant's lateralization index in correct and incorrect trials (for details, see Methods). Cosine amplitude quantifies the amplitude of rhythmic fluctuations of the lateralization index, which was significantly enhanced in correct compared to incorrect trials ($p = 0.003$) at frequencies 0.5–0.7 Hz (highlighted with light grey outline) around the digit presentation rate of 0.67 Hz. Shaded areas indicate ±1 SEM. Error bar in the inset indicates the 95-% confidence interval for the difference in cosine amplitude between correct and incorrect trials. (D) Scatterplot shows the average number of errors in incorrect trials as a function of average cosine amplitude at the 0.67-Hz digit presentation rate (averaged across 0.5–0.7 Hz) in incorrect trials for 19 participants (Spearman correlation; $r = –0.59$; $p = 0.009$).

The circular distance between cosine phases of correct and incorrect trials (averaged across frequencies 0.5–0.7 Hz) was significantly non-uniform (Rayleigh test; $z = 3.68$; $p = 0.023$; $r = 0.44$).

This indicates that the 0.67-Hz fluctuations of the alpha lateralization index in correct trials (average phase; ϕ = 0.94 radians) lagged behind these fluctuations in incorrect trials (average phase; ϕ = 1.04 radians). Note however that the average cosine phase was close to one (in correct and incorrect trials), indicating that peaks of the lateralization index were centered at individual digits, whereas troughs of the lateralization index co-occurred with time intervals in-between digits.

In order to relate the 0.67-Hz fluctuations of the lateralization index to a more fine-grained measure of performance, we calculated the Spearman correlation of participants' cosine amplitude (averaged across 0.5–0.7 Hz) in incorrect trials and the average number of errors in these incorrect trials (Figure 5.5D). The correlation showed a significant negative relationship (r = –0.59; p = 0.009), indicating that participants with a higher 0.67-Hz cosine amplitude in the lateralization index in incorrect trials made fewer errors.

5.3.5 Granger causal influence

Figure 5.6A shows average inter trial phase coherence (ITPC), alpha power, and alpha lateralization index across both experimental conditions (attention left/right) and 19 participants in the stimulation period (2.3–7.9 s). Each one of these three neural measures exhibited regular fluctuations with peaks centered at digit onset (for ITPC and alpha power) or at the mid-point of spoken digits (for the lateralization index). This lag called for a granger causal relationship analysis.

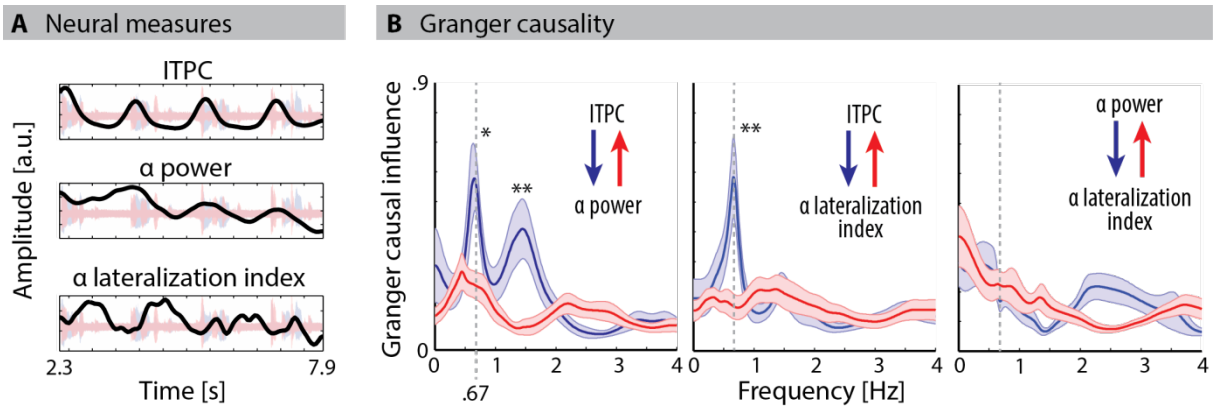

Figure 5.6. Granger causal influence. (A) Grand average of three neural measures in the stimulation period (2.3–7.9 s): inter trial phase coherence (ITPC), alpha power, and alpha lateralization index across all experimental trials and 19 participants. (B) Mutual granger causal influence for all pairs of neural measures at frequencies 0–4 Hz. Dashed vertical lines indicates the digit presentation rate (0.67 Hz). Shaded areas indicate ±1 SEM. * p < 0.05; ** p < 0.01.

The granger causal influence of ITPC on alpha power was larger than the granger causal influence in the reverse direction at 0.67 Hz (Wilcoxon signed-rank test; z = 2.5; p = 0.013; r = 0.57) and at 1.43 Hz (z = 2.66; p = 0.008; r = 0.61; Figure 5.6B left panel). Similarly, the granger causal influence of ITPC on alpha lateralization index at 0.67 Hz was significantly larger than the granger causal influence in the

reverse direction ($z = 2.66$; $p = 0.008$; $r = 0.61$; Figure 5.6B middle panel). The mutual granger causal influence between alpha power and alpha power lateralization at 0.67 Hz did not differ significantly ($z = 0.6$; $p = 0.546$; $r = 0.14$).

5.4 Discussion

In a dichotic listening task, we tested whether alpha power lateralization in the MEG indicates listeners' direction of attention to one of two ongoing speech streams. Results can be summarized as follows: First, in contrast to prior studies which focused on the anticipation of upcoming stimuli, we found alpha power lateralization also while participants were listening to ongoing speech. Second, source analysis revealed that alpha power modulations in parietal cortex, but critically also in auditory cortex regions underlie the attentional selection of one speech stream in a two-talker situation. Third, alpha power lateralization was not constant during selective attention but temporally aligned with the word rate.

5.4.1 Alpha lateralization in an attention-demanding dichotic listening task

Our behavioural results show that participants confused spoken digits on the ignored side with digits on the attended side more often than they reported digits presented on neither side (Figure 5.2A&B). This demonstrates strong competition of the two speech streams for attention which implies the need for selection of task-relevant and suppression of task-irrelevant speech, i.e., selective attention.

On the neural level, we observed a lateralization of alpha power while participants anticipated task-relevant speech on the left or right side (Figure 5.3B, anticipation period), which is in agreement with prior work across sensory modalities (e.g., Haegens et al., 2011a; Bauer et al., 2012; Ahveninen et al., 2013). Critically, alpha power was also lateralized while participants attended to ongoing speech on one side and ignored speech on the other side (for similar results in an EEG study, see Kerlin et al., 2010). Our source localization revealed both parietal as well as auditory cortex regions as neural sources of this alpha power lateralization during dichotic listening (Figure 5.4, stimulation period). It has been proposed that alpha power impacts neural processing of stimuli through a reduction of sensitivity in areas exhibiting high alpha power (for review, see Jensen and Mazaheri, 2010). Similar to vision and somatosensation, anatomical connections in the auditory system are predominantly contralaterally organized (e.g., Rosenzweig, 1951; Tervaniemi and Hugdahl, 2003). Thus, left auditory cortex is relatively more involved in processing sounds from the right ear and vice versa for right auditory cortex. Our source localization results suggest that sensitivity of auditory cortex regions is

enhanced (i.e., low alpha power) for task-relevant speech presented on the contralateral side, whereas sensitivity is reduced (i.e., high alpha power) for task-irrelevant speech on the contralateral side.

Notably, alpha power lateralization was stronger in right compared to left auditory cortex regions during dichotic listening which is in agreement with prior work (Müller and Weisz, 2012; Weisz et al., 2014) and possibly reflects that right auditory cortex is involved in processing sounds in the whole space, whereas left auditory cortex is mainly involved in processing sounds in the right space (Zatorre and Penhune, 2001). In prior studies, attentional modulation of alpha power in auditory cortex regions was exclusively observed during the anticipation of upcoming sounds (Müller and Weisz, 2012; Frey et al., 2014; Weisz et al., 2014). Thus, our results provide the first demonstration of alpha power modulation in auditory cortex regions as an underlying neural mechanism for spatial selective attention to ongoing speech.

Apart from auditory cortex regions, we found that parietal cortex regions contributed significantly to the alpha power lateralization during the entire trial (Figure 5.4; cue, anticipation, stimulation). Parietal cortex is part of a "dorsal attention network" (e.g., Sadaghiani et al., 2010), involved in orienting supramodal attention in space (Smith et al., 2010; Banerjee et al., 2011). Patient studies have shown that lesions of the inferior parietal lobe can cause distortions in the awareness of the contralateral space across sensory modalities (for reviews, see Driver and Mattingley, 1998; Vallar, 1998; for similar findings after temporal lobe lesions, see Karnath et al., 2001). In sum, we presume that the observed alpha power modulation in parietal cortex regions reflects orienting of attention to the left or right side of space, whereas alpha modulation in auditory cortex regions implements an attentional filter mechanism to suppress processing of task-irrelevant speech and to facilitate processing of task-relevant speech (Strauß et al., 2014).

5.4.2 Alpha lateralization temporally aligns with the input rate

Our results show that participants' deployment of high alpha power contralateral to task-irrelevant speech and low alpha power contralateral to task-relevant speech (alpha lateralization index; Figure 5.5A–C) was not constant during selective attention to ongoing speech. Instead, it fluctuated at a rate of 0.5–0.7 Hz, close to the digit presentation rate of 0.67 Hz. This is in agreement with *dynamic attending theory*, which states that attention fluctuates in synchrony with regular sensory stimulation (Large and Jones, 1999). Critically, strong alpha power lateralization – which presumably indicates strong selective attention – co-occurred with spoken digits, whereas alpha power lateralization was

reduced in uninformative time periods in-between digits. It has been shown that neural oscillations align with the external stimulation so that most critical stimulus segments fall into phases of high neural excitability (e.g., Cravo et al., 2013; Lakatos et al., 2013a; Wilsch et al., 2015). Extending these findings, we demonstrate here that also a well-studied neural signature of spatial attention (i.e., alpha power lateralization) aligns with the inherent regularity of ongoing speech (i.e., the word rate).

Importantly, our results reveal that the alignment of alpha power lateralization with the external stimulation is functionally significant for spatial selective attention. First, the modulation of lateralized alpha power at the word rate was larger in correct compared to incorrect trials (Figure 5.5C). This indicates that a higher intra-individual modulation of alpha power lateralization is associated with more successful spatial selective attention to speech. Second, participants with a larger modulation of alpha lateralization in incorrect trials made fewer errors in these incorrect trials (Figure 5.5D), showing that the alignment of alpha power lateralization with the speech signal explains inter-individual differences in selective attention. These results demonstrate that lateralized alpha power is a key neural signature for arguably one of the most relevant cognitive capabilities for speech processing in complex environments, i.e., selective attention. More generally, our findings support the often-proposed significance of alpha oscillations for effortful speech processing (Weisz et al., 2011; Obleser and Weisz, 2012; Becker et al., 2013; Wöstmann et al., 2015).

In agreement with one prior study (Kerlin et al., 2010), we found that alpha power lateralization was higher in the beginning of selective attention to one of two competing speech streams but decreased toward the end of a trial (Figure 5.5A). Considering our presumption that a larger alpha power lateralization indicates stronger spatial selective attention (see above), the decrease in alpha lateralization for later presented digits appears sub-optimal. However, we conjecture that alpha power lateralization is particularly involved in shifting attention in space. After attention has been shifted to task-relevant speech in the beginning of dichotic listening, alpha lateralization thus decreases (cp. Kerlin et al., 2010). This interpretation is in agreement with fMRI evidence for greater activity parietal cortex regions – which were among the sources of our alpha power lateralization – for switching compared to maintaining auditory spatial attention (Shomstein and Yantis, 2006).

5.4.3 Alpha lateralization – Driven by temporal expectation or external stimulation?

A question arising from our findings is whether regular fluctuations of the alpha lateralization index reflect participants' temporal expectation of upcoming digits or whether the alpha lateralization index

is driven by the regular external acoustic stimulation. To provide a definite answer to this question, our experimental paradigm would have to be changed, for instance by including trials with an irregular digit presentation rate. If the alpha lateralization index would then nevertheless align with the word rate this would indicate that alpha power lateralization can be driven by the external stimulation alone, which cannot be temporally expected in irregular trials.

One prior study in the somatosensory modality found increased lateralization of beta (15–30 Hz) but not alpha power at expected time points of stimulus presentation (van Ede et al., 2011). This suggests that neural oscillatory power lateralization can reflect participants' temporal expectations in the absence of external stimulation. To the contrary, our analysis of granger causality suggests that alpha lateralization was driven by the regular external stimulation. While granger causality is commonly used to investigate directed functional connectivity between different brain regions (e.g., Bosman et al., 2012), we used this measure to test temporal interdependencies between ITPC, alpha power, and alpha lateralization index. The time course of the neural encoding of spoken digits (quantified by ITPC) was highly predictive of the 0.67-Hz fluctuations of the alpha power lateralization index (Figure 5.6), suggesting that the encoding of each pair of simultaneously presented digits was followed by the selection of the digit on the attended side. Moreover, ITPC granger-caused fluctuations of alpha power at 0.67 Hz (and also 1.43 Hz), indicating that the neural encoding of the external stimulation was predictive of alpha power dynamics more generally. However, granger causality does not prove the causal dependence of underlying time courses. Thus, it might also be that fluctuations of the alpha lateralization would occur in the absence of any external stimulation when participants anticipate the presentation of digits. Future studies could manipulate external stimulation and temporal expectation independently to resolve this issue.

5.4.4 Conclusions

Our results show that the lateralized power of alpha oscillations in a parietal attention network and in auditory cortex regions is informative of spatial selective attention to ongoing speech in a complex listening situation. Stronger alignment of alpha lateralization with the word rate predicts more successful attentional selection. In sum, alpha power lateralization is an important neural marker to understand one of the most abundant cognitive challenges in everyday life, i.e., the attentional selection of signals in noise.

6 Study 4: Acoustic detail but not predictability affects distraction from irrelevant speech

This study describes a behavioural study of younger and older listeners' memory for task-relevant speech items under distraction from task-irrelevant speech. We asked in how far acoustic detail and the predictability of task-irrelevant speech determine how much it draws attention away from task-relevant speech.

6.1 Introduction

In Studies 1–3, we investigated the neural dynamics of selective attention to task-relevant speech. But to what degree is task-irrelevant speech drawing listeners' attention? And more important, does the attentional capture of irrelevant speech impede attention to task-relevant target speech? Study 3 provides initial hints to these questions: Participants confused irrelevant speech items with target speech more often than chance would predict. This suggests that irrelevant speech draws attention such that the semantic content (i.e., numerical values of spoken digits) captures attention at least on some trials, which interferes with attention to target speech. In the present study, we tested whether selective attention to target speech items in memory is affected by acoustic detail and predictability of irrelevant speech in younger and older listeners.

Imagine someone is asked to keep a nine-digit telephone number in mind for a short while. To minimize the risk of forgetting the serial order of digits, the person will automatically start to rehearse the digits in mind. This internal articulatory rehearsal is thought to be implemented in the *phonological loop*, which is a sub-system of working memory (Baddeley and Hitch, 1974; Baddeley, 1992). In the absence of distraction during encoding and rehearsal of digits, the person will possibly succeed at recalling the correct telephone number after a short while. However, what happens if the person simultaneously overhears a nearby conversation? Does the irrelevant speech necessarily draw attention away from the rehearsal of digits and impede attention to digits in memory? And in how far does this depend on acoustic and semantic properties of the irrelevant speech?

This particular situation is implemented in the *irrelevant speech task* (e.g., Colle and Welsh, 1976; Baddeley and Salame, 1986; Jones and Morris, 1992). Participants are (visually or acoustically) presented with a number of items (e.g., digits) they have to keep in memory in serial order. After a

retention period, they have to report the items in serial order. The first findings from this paradigm constituted the by now well-known *irrelevant speech effect*: The presentation of irrelevant speech during the encoding or retention impedes serial recall more than silence or white noise (e.g., Salame and Baddeley, 1987). This finding suggested that irrelevant speech disrupts the internal articulation of target items in the phonological loop. Similar results were also found for the presentation of irrelevant sounds (e.g., Jones and Macken, 1993; LeCompte et al., 1997; Tremblay et al., 2001), suggesting that the phenomenon is not specific to speech materials but rather depends on the presentation of varying acoustic items.

There is some evidence from prior research that acoustic as well as certain semantic features of irrelevant speech might affect the degree of interference. Thus, serial recall of target speech improves if the intelligibility of irrelevant speech is lowered through acoustic degradation using sine-wave speech (Tremblay et al., 2000) or noise-vocoding (Ellermeier et al., 2012). Moreover, serial recall is improved if semantics of irrelevant speech is reduced by means of presenting speech from a foreign language (Ellermeier et al., 2012) or random word lists (Tun et al., 2002). A shortcoming of the semantic manipulations used in prior studies is their low ecological validity since irrelevant speech in everyday listening situations is typically from the listener's native language and linguistically well-formed (i.e., not composed of random word strings). In the present study, we combined a well-studied acoustic manipulation (noise-vocoding) with an ecologically valid semantic manipulation not tested in the irrelevant speech paradigm before, that is, predictability of irrelevant speech. We tested the hypothesis that predictable irrelevant speech would be more distracting than unpredictable irrelevant speech. If, however, predictability of the unattended irrelevant speech materials would not affect interference, this would suggest that predictability requires listeners' attention in order to be processed.

One particularly interesting test case for this paradigm is the ageing listener. Cognitive capabilities generally decline at an older age (Park et al., 2003), importantly also the ability to ignore task-irrelevant sounds (Chao and Knight, 1997). Processing speech in noise requires attentional control, that is, the focus of attention on target speech and ignorance of noise. Older listeners show deficits in attentional control particularly if the noise is perceptually more salient than the signal (Passow et al., 2012). Research has shown that the semantic content of irrelevant speech affects performance stronger in older than younger listeners (Tun et al., 2002; Bell et al., 2008), suggesting that older listeners' attention is more likely drawn by semantic aspects of the irrelevant speech (for similar findings of visual distraction in a reading task, see Carlson et al., 1995).

We tested younger and older participants' serial recall of spoken digits under distraction from irrelevant speech varying in acoustic detail and predictability. In line with prior research, participants' recall of digits improved under stronger acoustic degradation of irrelevant speech. Predictability of irrelevant speech did not affect performance, suggesting that higher predictability does not increase the attentional capture of irrelevant speech.

6.2 Methods

6.2.1 Participants

12 younger (mean age: 27.1 years; age range: 23–33; 5 females) and 10 older participants (mean age: 67.9 years; age range: 61–78; 7 females) took part in this study. All participants were right-handed. They gave written consent and were financially compensated for participation.

6.2.2 Speech materials

Target speech comprised German digits from 1 to 9, spoken by a trained female speaker. Digits were recorded at a sampling rate of 44.1 kHz and were on average 0.6 sec long. All digits were equalized to the same root mean squared (RMS) amplitude of –30 dB full scale (FS). We generated 180 digit streams by concatenating the nine digits in random order with an onset-to-onset delay of 0.75 sec. The average length of digit streams was 6.6 sec (Figure 6.1A).

Irrelevant speech materials were adopted from a German version of the speech in noise (SPIN) sentences ("GSPIN", Erb et al., 2012), which were generated in a similar way as the original English SPIN sentences (Kalikow et al., 1977). All GSPIN sentences were composed of five to eight words and nine to eleven syllables. GSPIN sentences were spoken by a trained female speaker and recorded at a sampling rate of 22.05 kHz. Predictability of the final word in the GSPIN sentences is either high or low. A complete list of GSPIN sentences, as well as more details on generation and predictability ratings for these sentences can be found in Erb et al. (2012). For the present study, we chose 90 pairs of GSPIN sentences. Each sentence pair ended on the same mono- or bisyllabic noun (e.g., "net") which had a high predictability context in one sentence (e.g., "Paul caught the fish in his net."; translated from German) and a low predictability context in the other sentence (e.g., "Paul was talking about the new net."). Average sentence length was 2.1 sec, and did not differ significantly between high predictability (2.14 sec) and low predictability sentences (2.12 sec; $t_{178} = 0.8$; $p = 0.43$; $r = 0.06$). For simplicity, we will refer to the task-irrelevant GSPIN sentences as *irrelevant speech* hereafter.

In order to manipulate the spectral detail of irrelevant speech, all GSPIN sentences were noise-vocoded using three numbers of frequency channels (2ch, 8ch, and 32ch). In detail, sentences were divided in 2, 8, or 32 logarithmically spaced frequency channels between 70 and 9000 Hz. The speech envelope was extracted in each frequency channel and used to modulate a carrier of random noise in the channel's frequency range (for further details, see Erb et al., 2012). Finally, the signal was summed over all frequency channels and the amplitude was equalized to the intensity of spoken digits (–30 dB FS). A lower number of channels results in a more severe spectral degradation while the temporal information remains largely intact. Thus, fewer frequency channels significantly lower the intelligibility of noise-vocoded speech (e.g., Shannon et al., 1995; Obleser and Weisz, 2012).

6.2.3 Procedure

Participants were instructed to attend and remember spoken digits and to ignore subsequently presented irrelevant speech. Moreover, they were instructed to silently rehearse digits after these were presented. Each trial started with the binaural presentation of a stream of nine spoken digits in random order (Figure 6.1A). Irrelevant speech comprising three GSPIN sentences was presented 0.5 sec after the offset of the last digit (Figure 6.1B). The three sentences had an onset-to-onset delay of 2.67 sec. On each trial, three GSPIN sentences of the same predictability (high or low) and the same vocoding level (2ch, 8ch, or 32ch) were presented. During the presentation of all speech stimuli (digits and irrelevant speech), participants saw a central fixation cross on a computer screen. After the offset of the irrelevant speech, participants were visually presented with an array of digits from 1 to 9, arranged in random order (Figure 6.1C). Participants used the left mouse button to select the digits in the order of presentation. To facilitate the choice of digits, individual digits disappeared from the array after they were selected. The next trial started after an additional mouse click.

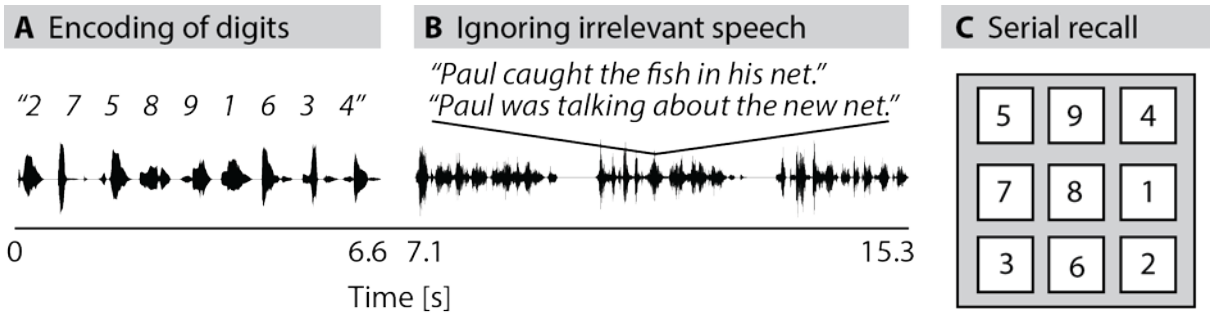

Figure 6.1. Design of the irrelevant speech task. (A) Participants encoded spoken digits from 1 to 9 presented in random order. The task was to retain the serial order of digits in memory during the presentation of irrelevant speech comprising three GSPIN sentences (B). Predictability of the final word in GSPIN sentences was either high (e.g., "Paul caught the fish in his net") or low (e.g., "Paul was talking about the new net"). GSPIN sentences were spectrally degraded, using noise-vocoding with three different numbers of frequency channels (2ch, 8ch, 32ch). (C) After acoustic stimulation, participants used the left mouse button to select the digits in the order of presentation from a visually presented array of randomly ordered digits.

Participants completed 120 trials, 20 for each condition in the 2 (predictability: high, low) × 3 (vocoding: 2ch, 8ch, 32ch) design. The order of trials was completely randomized across participants. Individual GSPIN sentences could occur more than once (at most three times) during the experiment, however not more than once during a single trial.

6.3 Results

Figure 6.2A shows average accuracy in the serial recall of digits for younger and older participants as a function of digit position, as well as predictability and acoustic detail (# of vocoder channels) of the irrelevant speech. Accuracies in the serial recall of digits were submitted to a repeated-measures ANOVA with the within-subject factors digit position (1–9), predictability (high, low) and acoustic detail of irrelevant speech (vocoding: 2ch, 8ch, 32ch). The between-subject factor was age group (younger, older). The main effect digit position was significant (Greenhouse-Geisser correction of p-value due to violation of sphericity, Mauchly's test: $p < 0.001$; $F(8, 160) = 67{,}54$; $p < 0.001$; $\eta^2_P = 0.77$). This indicates that the accuracy for the recall of digits significantly varied over digit positions.

6.3.1 Acoustic detail affects task performance

The main effect acoustic detail was significant ($F(2, 40) = 25.61$; $p < 0.001$; $\eta^2_P = 0.56$). Post-hoc paired t-tests revealed that accuracy was significantly higher for 2ch compared to 8ch ($t_{21} = 5.25$; $p < 0.001$; $r = 0.75$), for 2ch compared to 32ch ($t_{21} = 6.38$; $p < 0.001$; $r = 0.81$), but not significantly different between 8ch and 32ch ($t_{21} = 1.62$; $p = 0.119$; $r = 0.33$). Thus, participants performed better when the irrelevant speech was more severely degraded. Moreover, the digit position × acoustic detail interaction was significant ($F(16, 320) = 3.08$; $p < 0.001$; $\eta^2_P = 0.13$), indicating that the benefit from more severe noise-vocoding of irrelevant speech was stronger for digits presented at later positions (Figure 6.2A).

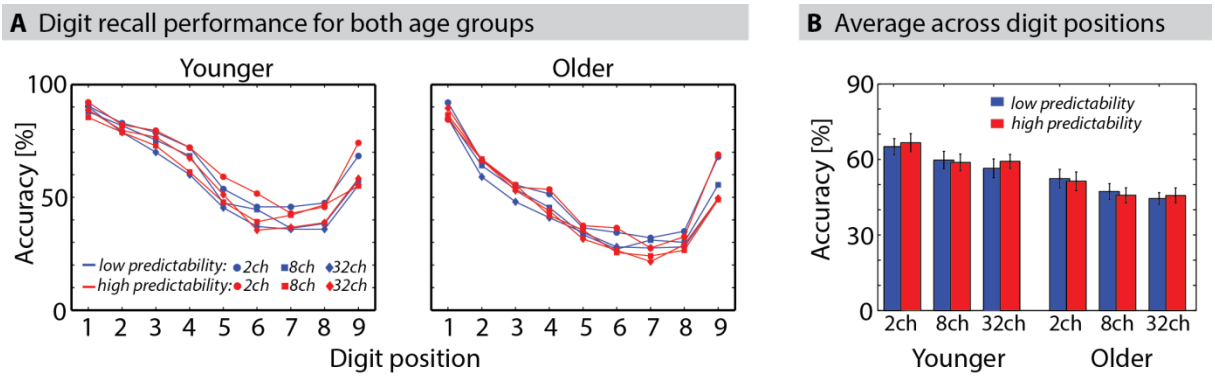

Figure 6.2. Accuracy in the irrelevant speech task. (A) Average accuracy of serial recall as a function of digit position for all conditions in the 2 (predictability: high, low) x 3 (vocoding: 2ch, 8ch, 32ch) design. Recall accuracy was highest for initial digits (digit positions 1–3) and the final digit in the stream (digit position 9) for younger and older participants. (B) Recall accuracy averaged across digit positions for younger and older participants. Recall accuracy decreased with more

acoustic detail (higher number of vocoder channels) of the irrelevant speech. Overall performance was reduced for older compared with younger participants. Error bars indicate ± 1 SEM.

6.3.2 Predictability of irrelevant speech does not impact performance

Neither the main effect predictability ($F(1, 20) = 0.18$; $p = 0.676$; $\eta^2_P < 0.01$), nor the predictability × acoustic detail interaction was significant ($F(2, 40) = 0.93$; $p = 0.401$; $\eta^2_P = 0.05$). To assess in how far these null-findings indicate the absence of an effect of predictability or whether our data were just insensitive in finding an effect, we calculated the Bayes Factor (using R studio version 0.97.551, and the *BayesFactor* package). When comparing two models, the Bayes Factor indicates how many times more likely the observed data are under the alternative compared to the null-model (Dienes, 2014). The Bayes Factor indicates support for the alternative model when it is larger than 3 and support for the null model when it is smaller than 0.33.

To compute the Bayes Factor for the main effect predictability, we compared the alternative model (random factor: participant, fixed factor: predictability) with the respective null-model (random factor: participant). The resulting Bayes Factor was < 0.2, indicating support for the null-model and suggesting that predictability had no effect in the present study. For the predictability × acoustic detail interaction, we compared the full model (random factor: participant, fixed factors: predictability, acoustic detail, predictability × acoustic detail) with the respective null-model containing the same factors except the interaction term. The resulting Bayes Factor was < 10^-8, providing substantial support for the null model and suggesting that the predictability × acoustic detail interaction had no effect in the present study.

For completeness, we also computed the Bayes Factor for the main effect acoustic detail (which was highly significant; see above) by comparing the alternative model (random factor: participant, fixed factor: acoustic detail) with the respective null-model (random factor: participant). The resulting Bayes Factor was > 10^7 indicating substantial support for an effect of acoustic detail on accuracy.

6.3.3 Task performance decreases with age

The main effect age group was significant ($F(1, 20) = 9{,}55$; $p = 0.006$; $\eta^2_P = 0.32$), indicating lower overall accuracy for older compared with younger participants. Moreover, the digit position × age group interaction was significant ($F(8, 160) = 2.09$; $p = 0.039$; ; $\eta^2_P = 0.1$). Post-hoc independent samples t-tests revealed that older participants showed a significantly lower accuracy than younger participants only at digit positions 2, 3, 4, 5, and 7 (all $p < 0.05$; all $r > 0.42$). None of the remaining interactions with age group were significant (all $p > 0.15$; all $\eta^2_P < 0.05$).

6.4 Discussion

We tested whether younger and older listeners' memory for spoken target digits would be affected by the attentional capture of irrelevant speech varying in acoustic detail and predictability. Results can be summarized as follows: (1) More acoustic detail of irrelevant speech decreased serial recall of target speech in younger and older listeners. (2) Predictability of irrelevant speech had no significant impact on performance. (3) Overall, younger listeners performed better than older listeners.

6.4.1 Distraction from irrelevant speech depends on acoustic detail

The main finding of the present study was a decrease in the serial recall performance of target speech items when irrelevant speech with more acoustic detail was presented during the retention of target speech in memory. This result is in agreement with prior research showing that serial recall of target speech is negatively affected if the intelligibility of irrelevant speech is enhanced (Tremblay et al., 2000; Ellermeier et al., 2012).

Noise-vocoding degrades spectral information in the acoustic signal while the temporal envelope, which is particularly important for speech recognition (Shannon et al., 1995), remains largely intact. Thus, if ≥ 8 frequency bands are used for noise-vocoding, speech in quiet is fairly intelligible for younger (Obleser et al., 2007; Obleser et al., 2008) as well as for older listeners (Sheldon et al., 2008). In the present study, irrelevant speech was unintelligibly only in the 2ch condition, but largely intelligible for 8ch and 32ch. It is likely that intelligibility of irrelevant speech was driving the significant effect of acoustic detail here since statistical analyses revealed that only the unintelligible 2ch condition differed significantly from the two intelligible conditions (8ch, 32ch). This is in line with results of one previous study where the strongest improvement in the recall of target speech was also observed when vocoder bands increased from 2ch (unintelligible) to 20ch (intelligible) (Ellermeier et al., 2012). In the present study, irrelevant speech was presented during the retention period when participants internally articulated the target speech items in the phonological loop of working memory (Baddeley and Hitch, 1974). Our results thus indicate that participants' rehearsal in the phonological loop was impaired by the presence of intelligible task-irrelevant speech.

One possible interpretation is that intelligible irrelevant speech draws participants' attention away from the internal rehearsal of target speech items. Research has shown that listeners recall a significant amount of irrelevant speech in a surprise recognition test (Tun et al., 2002), indicating that irrelevant speech cannot be ignored entirely. Successful internal rehearsal of target speech requires that attention

is focused on target speech items in working memory. If attention is drawn to irrelevant speech, it might partially occupy the limited capacity of working memory (Miller, 1956; Simon, 1974; Cowan, 2001), which then impedes memory for target speech (for a review on interactions between attention and working memory, see Awh et al., 2006). This interpretation is also supported by neuroimaging data (Gisselgard et al., 2004) showing that irrelevant speech modulates activity in dorsolateral prefrontal cortex similarly to verbal working memory load.

6.4.2 Predictability of irrelevant speech does not affect memory for target speech

One unexpected result of the present study was that predictability of irrelevant speech did not significantly impact younger and older participants' serial recall of target speech items. The observed small Bayes Factors (< 0.3) for the main effect predictability and the interaction with acoustic detail support the absence of an effect of predictability rather than indicating insensitivity of the data (for detailed information on the statistical approach, see Kruschke, 2011). This finding somewhat contradicts previous studies, where the semantic content of irrelevant speech affected older participants' memory for target stimuli (Tun et al., 2002; Bell et al., 2008). The irrelevant speech materials from the present study were used previously as task-relevant speech, where predictability significantly improved speech comprehension under noise-vocoding with 4ch (Hartwigsen et al., 2014) and 8ch (for similar speech materials in English, Obleser et al., 2007). Thus, we conclude that predictability of speech requires attention to affect performance. Since participants in the present study were instructed to ignore the irrelevant speech, predictability did not affect task performance.

6.4.3 Older listeners show a decreased memory for target speech

Although the overall pattern of results was similar for both age groups (Figure 6.2), older listeners performed generally worse compared with younger listeners. Memory capacity typically shows a decrease with age (Fisk and Warr, 1996; Baltes and Lindenberger, 1997; Nilsson, 2003), which likely affects older participants' ability to retain nine items in serial order in memory. Moreover, hearing acuity is generally reduced in older listeners (Frisina, 2009). Reduced hearing acuity likely leads to a higher demand of encoding target speech, which causes additional load in working memory (see Pichora-Fuller et al., 1995; Wingfield et al., 2005; Lunner et al., 2009). Thus, a general decline in older listeners' performance in the irrelevant speech task was a highly expected result.

6.4.4 Conclusions

Our results suggest that the attentional capture of irrelevant speech increases with higher speech intelligibility. Contrary, predictability of speech requires attention to be processed and thus did not affect distraction from ignored irrelevant speech. While memory for speech generally declines at an older age, the degree of distraction from more intelligible or more predictable irrelevant speech does not increase at an older age. Future studies could use electrophysiological measures described in previous chapters (e.g., alpha power dynamics, neural phase-locking) to foster an understanding of the underlying brain mechanisms of auditory distraction in the irrelevant speech paradigm (Kopp et al., 2004; Schlittmeier et al., 2011).

7 General Discussion

The present thesis comprises four studies which investigate the behavioural and neural dynamics of selective attention to speech under demanding listening conditions. A major objective of this thesis is to develop an initial framework to explain in how far mechanisms of selective attention might counteract acoustic degradation to support speech comprehension (see section 7.6). Since this research goal requires also to understand to what extent attention mechanisms are limited in populations that experience particular difficulties in attention-demanding situations, three studies of this thesis compare younger with older listeners. Detailed discussions of all experimental results can be found in the discussion sections of Studies 1-4. In the present chapter, I will adopt a broader perspective to integrate results of individual studies into a common framework of the neural dynamics of selective attention to speech in demanding listening situations. The following section summarizes the major results and their interpretations as answers to the research questions in section 1.3.

7.1 Summary of experimental results

The present thesis started out with an electroencephalography (EEG) study of the event-related potential (ERP) in response to degradation of the temporal fine structure in speech materials. Study 1 revealed that acoustic degradation guides listeners' allocation of attention in an auditory number comparison task. In detail, larger amplitudes of the contingent negative variation (CNV) indicated that the allocation of attention increased when listeners were faced with more degraded acoustics. The reliability of this result was approved by replication in a different sample of participants in Study 1.2. Important for the framework proposed below (see section 7.6), speech comprehension decreased with acoustic degradation but increased with the amplitude of the CNV, showing that neural mechanisms of attention counteract acoustic degradation. Considering age effects, the acoustics-driven modulation of the CNV was preserved in older listeners, speaking for a robust allocation of attention, at least when overall acoustic conditions are carefully adapted to older listeners' requirements. In a follow-up experiment, Study 1.2 manipulated task-relevance by offering monetary incentives on each trial. Increasing incentives from 1 to 5 cent did not affect CNV amplitude but the CNV was larger in this follow-up experiment compared to the main experiment where no incentives were offered (Study 1.1). Larger CNV amplitude thus speaks for the enhanced allocation of attention in a context of high task-relevance.

In Study 2.1, we analysed brain oscillatory mechanisms in the same dataset used in Study 1.1. Study 2.1 found that increased acoustic detail but also better predictions (i.e., of the numerical size of the second digit in the auditory number comparison task) led to reduced power of neural alpha oscillations. Interestingly, acoustic detail affected behavioural responses and neural alpha power dynamics stronger in older listeners, indicating that attention to speech in the presence of distraction is particularly dependent on acoustic conditions at an older age. Across age groups, stronger alpha power modulation predicted lower subjective difficulty in everyday listening situations, suggesting that neural attention mechanisms compensate for acoustic challenges.

In an additional analysis of the same dataset, Study 2.2 showed that listeners' neural responses exhibited a characteristically different neural phase-locking to the temporal envelopes of attended speech (i.e., spoken digits) vs unattended speech (i.e., background masker). Thus, attention modulates neural responses to basic acoustic features, possibly to segregate attended and unattended speech signals. Reliability of results from Studies 2.1 and 2.2 was further established by replication of the prediction-effect on neural alpha power as well as the neural phase-locking to attended and unattended speech in a similar paradigm in a different sample of older listeners (unpublished data; analysed by Dunja Kunke from the research group "Auditory Cognition" at the MPI in Leipzig, Germany).

Study 3 investigated in a magnetoencephalography (MEG) experiment in how far the power of neural alpha oscillations reflects which auditory stream is in the listeners' current focus of attention. In a dichotic listening paradigm, Study 3 found that recording sites over the hemisphere ipsilateral to the attended speech signal show a relative increase of alpha power, whereas contralateral recording sites show a relative decrease of alpha power. The sources of this alpha power lateralization were found in parietal, frontal, and auditory cortex regions. Considering the functional inhibition framework of alpha oscillations (Jensen and Mazaheri, 2010), these results suggest that neural activity in supramodal (i.e., parietal) and primary sensory (i.e., auditory cortex) regions is regulated to attain spatial selective attention. In other words, these findings indicate that alpha power modulations implement an *attentional filter mechanism* that enhances neural processing of the signal and suppresses processing of the noise. Interestingly, alpha power lateralization was not constant during selective attention to ongoing speech but fluctuated at the word rate. Larger amplitudes of these fluctuations predicted better recall of attended speech items, suggesting that fluctuating alpha power lateralization at the word rate supports selective attention to speech in noise.

Finally, Study 4 explored in a behavioural experiment which features of an irrelevant speech signal determine how much it interferes with memory for target speech. Using an *irrelevant speech paradigm*, Study 4 found that more acoustic detail (i.e., a larger number of vocoder bands) but not higher predictability of irrelevant speech impairs task performance. Although overall performance decreased in older listeners, the relative increase of distraction from task-irrelevant speech with more acoustic detail was unchanged at an older age. These results indicate that the degree to which irrelevant speech captures attention increases with higher intelligibility through more preserved acoustic detail. To the contrary, predictable irrelevant speech does not capture attention more than unpredictable irrelevant speech.

Taken together, the four studies of this thesis revealed behavioural and neural signatures of selective attention to speech in noise. In the following section, I will integrate the observed neural signatures to draw a more coherent picture of how neural dynamics reflect the attentional selection of relevant signals despite noise.

7.2 Integrating the various neural signatures of selective attention

In the present thesis, we found three neural signatures reflecting demands on attention; namely the contingent negative variation (CNV, Study 1), the power of neural alpha oscillations (Studies 2.1&3), and the neural phase-locking to the temporal envelope of speech (quantified by the cross-correlation method, Study 2.2). But in how far are these neural signatures interrelated or do even reflect the same underlying brain process? A first approach to answer this question is to test for statistical dependencies among these neural signatures. Studies 1.1, 2.1, and 2.2 are all based on the same dataset of eighteen younger and twenty older participants. Thus, I conducted a post-hoc analysis to test for correlations among neural signatures observed in these studies. In detail, I extracted the linear coefficients (i.e., slopes) quantifying the effect of increasing acoustic detail on (1) CNV magnitude (in the significant cluster in Study 1.1), (2) alpha power (averaged across clusters A1 and A2 in Study 2.1), and (3) neural phase-locking to the envelope of attended speech (averaged across cross-correlation time-lags 110–230 ms in Study 2.2). There was no significant relation among linear coefficients from different studies (pairwise linear Pearson's correlations; all $r < 0.21$; all $p > 0.2$). Although this analysis does not prove that these neural signatures are entirely independent, it at least suggests that CNV, alpha power, and neural phase-locking do not reflect the very same underlying brain process. Thus, a more nuanced discussion of these neural signatures is required, which follows below.

An important dimension on which CNV, alpha power, and neural phase-locking diverge is their temporal occurrence. Study 1 has shown that the CNV is triggered by the onset of the distracting speech masker. With more severe acoustic degradation, the CNV increases already before task-relevant digits are presented in the auditory number comparison task. Thus, we interpreted larger CNV magnitude to reflect increased allocation of attention in anticipation of higher listening demands. This is in agreement with prior studies that found task-related modulations of the CNV prior to the onset of target stimuli (e.g., Rebert et al., 1967; McCallum and Walter, 1968; Rockstroh et al., 1993). Thus, the most critical time period for the CNV is before the onset of (speech) stimuli. It is in this time period where the CNV indicates the degree of attention allocation. To the contrary, neural phase-locking to the temporal envelope of speech cannot take place in anticipation but only in response to a speech stimulus. Study 2.2 has shown that the EEG signal differentially phase-locks to the temporal envelope of attended and unattended speech with a time-lag of ~170 ms. Obviously, the attentional selection of an auditory object cannot take place before the object is presented. Thus, in contrast to the CNV which reflects the allocation of attention prior to target speech onset, neural phase-locking reflects whether an ongoing speech signal is attentionally selected (with a temporal delay commonly ranging between 100 and 200 ms; see also Ding and Simon, 2012; Zion Golumbic et al., 2013; Hambrook and Tata, 2014; Kong et al., 2014).

The CNV and neural phase-locking are measures of evoked activity, which means that they are strictly time- and phase-locked to the acoustic stimulation (for a more detailed comparison of evoked and induced activity, see section 2.3.2). In contrast, induced alpha oscillations are not necessarily time-locked to sensory events. This becomes most obvious in memory paradigms, where alpha power modulations are observed in a delay interval that is free of any sensory stimulation (e.g., Jensen et al., 2002; Sauseng et al., 2009; Obleser et al., 2012). In Study 2.1, alpha power modulations were observed during the encoding of task-relevant digits and also in the absence of task-relevant speech during the prediction of the second digit. In Study 3, alpha power modulations were observed both during the anticipation and during the actual presentation of speech stimuli. These observations demonstrate that opposed to the CNV and neural phase-locking, alpha power modulations can be detached from the external acoustic stimulation. Based on this and other observations (for direct neural evidence, see e.g., Buffalo et al., 2011; van Kerkoerle et al., 2014), it has been proposed that alpha oscillations do not directly reflect perceptual processing, but rather a top-down mechanism to regulate the processing of relevant and irrelevant information (e.g., Jensen et al., 2012; Klimesch, 2012).

But what are the underlying brain physiological mechanisms of these three neural signatures of selective attention? Generally, fluctuations in the M/EEG signal arise from changes in the excitability of the underlying neural tissue (e.g., Buzsaki and Draguhn, 2004; Lakatos et al., 2005). Regarding the CNV, it has been proposed that slow cortical potentials of negative polarity are likely generated by the depolarization of cortical pyramidal cells (Raichle, 2011), indicating a state of enhanced neuronal excitability. Thus, a larger CNV possibly indicates enhanced excitability, which might support selective attention to barely perceptible stimuli. Following this line of argumentation, phase-locking of fluctuations in the M/EEG signal to the speech envelope reflects the alignment of phases of high and low excitability to the temporal structure of speech. In particular, neural phase-locking to the speech signal might indicate that phases of high excitability align to critical parts of the attended speech signal, whereas phases of low excitability align to acoustic events of unattended speech (Ahissar et al., 2001; Nourski et al., 2009; Giraud and Poeppel, 2012). This interpretation is in agreement with studies showing that neural oscillations align with the external stimulation so that most critical stimulus segments fall into phases of high excitability (e.g., Henry and Obleser, 2012; Cravo et al., 2013; Lakatos et al., 2013a; Wilsch et al., 2015). Finally, in how far does alpha power capture fluctuations in neuronal excitability? It has been shown that neuronal excitability varies over the course of an alpha cycle (e.g., Dugue et al., 2011). Excitability is highest in the trough of an alpha cycle, evidenced by increased neuron firing rates (e.g., Haegens et al., 2011b) and improved stimulus detection (e.g., Busch et al., 2009; Strauß et al., 2015). If oscillations at the alpha frequency are shallow (i.e., low alpha power) the high-excitability phases are prolonged. To the contrary, large amplitudes of alpha oscillations (i.e., high alpha power) reflect reduced high-excitability phases and thus decreased neural activity (e.g., Klimesch et al., 2007; Haegens et al., 2011b). In consequence, alpha power has been proposed to support selective attention, through inhibition (vs enhancement) of activity in brain regions that are task-irrelevant (vs task-relevant) (Jensen and Mazaheri, 2010; Foxe and Snyder, 2011; Klimesch, 2012).

Taken together, CNV, alpha power, and neural phase-locking reflect temporally and functionally distinct mechanisms of attention. The CNV is triggered before the onset of task-relevant speech and reflects the anticipatory allocation of attention. Neural phase-locking follows the speech signal with a time-lag of a few hundred milliseconds and segregates attended and unattended speech on a neural level. Alpha power is not necessarily temporally bound to the sensory stimulation and reflects the inhibition of task-irrelevant brain areas and processes.

7.3 Revisiting early vs late selection

All four studies in the present thesis used experimental paradigms requiring participants' to comprehend one speech signal despite acoustic distraction (i.e., noise). One conjecture of this thesis is that selective attention supports speech comprehension, presumably through signal enhancement and noise suppression. To understand how attention selects task-relevant speech from the plethora of sounds arriving at the ear, it is crucial to consider at which level of processing the noise is filtered out and the target speech signal is selected for further processing. Thus, in this section, I will integrate the results from this thesis into the early and late selection theories of attention (see also section 1.1.1). Early selection theory states that the unattended signal is filtered out on the basis of fundamental physical features such as pitch or location (Broadbent, 1958). With respect to brain imaging, it follows from early selection theory that attention modulates neural responses to basic sensory features (e.g., temporal structure) at "early" processing stages (e.g., in auditory cortex). To the contrary, late selection theory holds that the unattended signal is filtered out after higher-level features such as the semantic information are extracted (Deutsch and Deutsch, 1963). Thus, it follows from late selection theory that attention modulates neural responses to higher-level features at "late" (i.e., non-primary) processing stages (Serences and Kastner, 2014).

Studies 1 and 2.1 do not lend unequivocal support to either early or late selection theory. In Study 1, larger amplitudes of the contingent negative variation (CNV) in the EEG indicated increased allocation of attention under more degraded acoustic conditions. This result demonstrates that basic acoustic features guide the preparatory allocation of attention in anticipation of a demanding listening task. However, this finding does not reveal whether the attentional selection of the signal is accomplished on the basis of fundamental or higher-level features. In Study 2.1, stronger modulations of neural alpha power indicated that acoustic detail and also predictions about upcoming speech content facilitated neural mechanisms of selective attention. Again, these results can however not provide any evidence whether selective attention is implemented at early or late processing stages.

Study 2.2 provides more explicit support for the early selection theory. The main result of this study was the differential phase-locking of the neural response to the temporal envelopes of attended and unattended speech (for similar results, see e.g., Ding and Simon, 2012; Hambrook and Tata, 2014; Kong et al., 2014). The temporal envelope constitutes a basic physical feature of the speech signal that can be extracted from the sound waveform (Smith et al., 2002). Findings of Study 2.2 thus demonstrate that attention modulates in how far basic acoustic features are processed neurally, which is in

agreement with early selection theory. Furthermore, also Study 3 supports early selection theory by showing that attention modulates the power of alpha oscillations at comparably early stages of processing, i.e., in auditory cortex regions. In detail, this study revealed that listeners' direction of spatial selective attention in a dichotic listening task modulates alpha power in parietal, frontal, and importantly also in auditory cortex regions. This is in agreement with evidence from fMRI studies that found attentional modulations of neural activity in early visual areas (V1, Gandhi et al., 1999; LGN, O'Connor et al., 2002). Results of Study 3 thus suggest that attention enhances acoustic features of the speech signal and suppresses acoustic features of the noise in auditory cortex regions, possibly to reduce noise-interference on subsequent processing stages (for a more nuanced discussion of these results, see section 5.4). Taken together, electrophysiological results from Studies 2.2&3 indicate that the attentional selection of target speech sets in at early processing stages where basic acoustic features (e.g., temporal envelope, spatial location) are analysed, which is in line with the early selection theory.

In addition to the electrophysiological evidence (see above), Study 4 provides behavioural support for the early selection theory. In this study, basic acoustic features (i.e., number of vocoder bands) but not semantic features (i.e., final-word predictability) of unattended speech had an impact on listeners' memory for attended speech. Basic acoustic features of the unattended signal affected task performance, showing that they were processes. In contrast, semantic features of the unattended signal had no impact on task performance, suggesting that unattended speech was "filtered out" before semantic features were processed. Although results of this thesis are in agreement with early rather than late selection theory, it is important that they do not show that late selection does not take place at all. It is likely that in different task settings, attentional modulation of neural activity would occur also at later processing stages (e.g., Vogel et al., 1998). Moreover, brain imaging findings of attentional modulations at early processing stages (e.g., in auditory cortex in Study 3) might be driven by feedback from later processing stages at which semantic features of attended and unattended signal are processed. Thus, it is conceivable that attentional selection of speech in noise is implemented at early and also late processing stages. However, based on our results, I consider it unlikely that selective attention to speech in noise is realized by late selection alone (see also Serences and Kastner, 2014).

7.4 Revisiting selective attention in younger and older listeners

Hearing acuity (Brant and Fozard, 1990) and cognitive capabilities (Park et al., 2003) show a negative trajectory with age, which might explain listening difficulties of the elderly (Humes, 1996; Pichora-

Fuller and Souza, 2003; Wingfield et al., 2005). In this regard, it might be surprising that this thesis found only few age-differences in behavioural and neural responses in speech comprehension tasks. In Studies 1.1 and 2, individual adjustments of stimulus materials systematically equalized younger and older listeners' overall task accuracy to the same level. In detail, stimulus intensity was adapted to participants' audiometric thresholds (Moore et al., 1998) and an adaptive tracking procedure was applied to estimate the signal to noise ratio (SNR) required to perform at ~71 % correct (Levitt, 1971). Important for the present thesis, older participants' individual acoustic adjustments were predictable from their cognitive capabilities (i.e., working memory capacity; see Figure 3.2). Older listeners with larger working memory capacities were able to perform the auditory number comparison task at lower SNRs. In line with other studies, this result demonstrates that speech comprehension in demanding listening conditions does not only depend on hearing acuity, but also on cognitive functioning (e.g., Pichora-Fuller, 2003a; Lunner et al., 2009; Neher et al., 2009). One interpretation is that working memory is required to fill in or to infer missing information of degraded acoustic input in order to match it to a phonological representation in long term memory (Rönnberg et al., 2013). In consequence, it is likely that our individual acoustic adjustments compensated not only for decreased hearing acuity but also for the decline in cognitive capabilities at an older age. This has two important implications: First, this might explain why this thesis found relatively few age-differences in behavioural and neural responses. Second, this demonstrates the tight interdependencies between cognitive functioning and acoustic input degradation, which are further elucidated on in the framework proposed below (see section 7.6).

But how are similarities and differences between age groups in Studies 1.1, 2, and 4 interpreted? Study 1.1 found that acoustic conditions guide the preparatory allocation of attention similarly in younger and older listeners. Across age, listeners increase the allocation of attention if degraded acoustics indicate high demands on attention for speech comprehension. Our control experiment (section 3.1.3.5) has shown that this allocation of attention is partly automatic and partly driven by expected task difficulty. Based on the same dataset, Study 2.2 found no age-effect on listeners' neural phase-locking to the temporal envelope speech. Preserved acoustic detail improved the segregation of the temporal envelopes of attended and unattended speech on a neural level. Taken together, results of Studies 1.1 and 2.2 suggest that the preparatory allocation of attention and the attentional modulation of neural phase-locking to speech can be preserved at an older age, given that acoustic conditions are carefully adapted to individual requirements.

In Study 2.1, older compared with younger listeners' demands on attention were reduced more strongly with increasing acoustic detail, evidenced through a larger behavioural benefit and a larger decrease in neural alpha power. But how substantial are these age-effects in comparison to other studies on differences between age groups in the area of neuropsychology? To arrive at an initial answer to this question, I compared effect sizes of age-effects in Study 2.1 to effect sizes of six studies from different groups, which also report significant two-way interactions with age group on behavioural and/or neuroimaging data (Tun et al., 2002; Gazzaley et al., 2007; Gazzaley et al., 2008; Zanto et al., 2010; Meister et al., 2012; Passow et al., 2014). Partial eta-squared effect sizes in Study 2.1 (ranging between 0.085 and 0.444) were comparable or even larger than effect sizes in these other studies (ranging between 0.0446 and 0.303). Partial eta-squared values of 0.01, 0.06 and 0.14 indicate small, medium, and large effects, respectively (Cohen, 1969). In sum, although this effect size comparison has by far not the status of a systematic meta-analysis, it at least suggests that the observed age-effects in Study 2.1 are of medium to large size and considerable in the field of neuropsychology.

Thus, age-effects in Study 2.1 require a thorough interpretation. Prior work has shown that older listeners experience difficulties in ignoring salient task-irrelevant sounds (Chao and Knight, 1997; Tun et al., 2002; Passow et al., 2012), which indicates a decline in attentional control at an older age. In this regard, findings of Study 2.1 might reflect that older listeners are generally more dependent on the external acoustic input, whereas younger listeners are less driven by acoustic features and relatively more by listening strategies such as the prediction of upcoming speech content. Anecdotal evidence for this interpretation comes also from our participants' subjective reports after accomplishing the auditory number comparison task (Studies 1&2). Most of the older participants expressed their interest in the audiobook which served as the task-irrelevant speech masker. In particular, they asked the experimenter about the content and narrator of the audiobook. In contrast, younger participants did not show any interest in the task-irrelevant audiobook. Although this observation provides no empirical evidence, it is in agreement with the interpretation that older listeners' focus of attention is particularly dependent on the acoustic input, which is dominated by the task-irrelevant audiobook in our auditory number comparison task. Interestingly, older listeners' dependence on the acoustic stimulation can lead to decrements or improvements of speech comprehension, depending on the task setting. If the acoustic stimulation draws attention to salient but task-irrelevant stimuli, task performance decreases at an older age (Passow et al., 2012; Passow et al., 2014). However, if salient acoustic cues such as preserved temporal fine structure support speech comprehension, task

performance increases particularly strong at an older age (Study 2.1). Future studies could test the same group of younger and older adults in both of these task settings to examine the validity of our conjecture that older listeners are more dependent on the acoustic stimulation.

In contrast to Studies 1.1 and 2, Study 4 did not include individual adjustments of acoustic conditions to younger and older listeners' requirements. In the irrelevant speech task, overall performance was thus lower for older compared with younger adults. This finding was likely due to reduced working memory capacity in older adults (e.g., Salthouse and Kersten, 1993; Fisk and Warr, 1996), which impairs the serial recall of digits in the irrelevant speech task. In addition, it might be that older listeners were more distracted from irrelevant speech, which interfered with attention to the mental rehearsal of digits. Critically, increases in performance with acoustic degradation were unaffected by listeners' age. Thus, although the general distraction from irrelevant speech might increase at an older age (Bell et al., 2008), acoustic and semantic features of irrelevant speech do not necessarily increase the degree of distraction in an age-specific manner (however, see Tun et al., 2002).

7.5 Limitations of the present research

It is important to note some limitations of the experimental work of this thesis. First, our analysis of age-effects was limited to the contrast of two groups of younger (20–30 years) and older adults (60–80 years) in Studies 1.1, 2, and 4. Thus, this thesis does not reveal the underlying trajectory of attention mechanisms as a function of age. Other studies found that attentional control increases from middle to late childhood (Passow et al., 2013) and decreases again from young to late adulthood (Passow et al., 2012), suggesting an inverted u-shape function of attentional control with age. However, in how far does hearing loss affect this change in attention mechanisms with age? Hearing loss often starts in the 4th decade of life (Bhatt et al., 2001; Jennings and Jones, 2001); an age group not examined in this thesis. Thus, this thesis does not reveal in how far age-related changes in neural dynamics are driven by hearing loss itself, hearing loss-induced structural and functional reorganization of the auditory and neural system (e.g., Syka, 2002; Tremblay and Ross, 2007; Peelle et al., 2011), cognitive decline (Park et al., 2003), or an interaction among these factors (Wingfield et al., 2005). Furthermore, younger and older listeners in this thesis were not matched for educational and socioeconomic status, which complicates the interpretation of the underlying mechanisms driving the observed age-effects. One desirable solution to these limitations would be the study of longitudinal changes of neural mechanisms for speech comprehension from middle to late adulthood.

Second, in Studies 1.1 and 2 we made a great effort to carefully adapt the listening conditions (i.e., overall stimulus intensity and SNR) to younger and older listeners' requirements. Adjustments were necessary from a practical point of view, since fixed acoustic conditions across participants would exceed the relatively narrow dynamic range in which acoustic stimuli modulate common accuracy measures in speech comprehension tasks. That is, if acoustic conditions are too poor or too favourable, listeners perform at chance level (i.e., 50 % correct in case of two response options) or at ceiling (i.e., 100 % correct), respectively. This would abolish effects of independent variables on response accuracy. However, the disadvantage of individual adjustments is that they might obscure existing age-effects that could be observed without these adjustments (see also Schneider et al., 2000). For instance, it might be that sensitivity to temporal fine structure decreases due to a decline of auditory acuity at an older age (e.g., Grose and Mamo, 2010; Lunner et al., 2012; Moore et al., 2012). If acoustic conditions are carefully adjusted to compensate for inter-individual differences in auditory acuity, it might be that also age-effects on sensitivity to temporal fine structure are reduced. In order to understand the effects of individual adjustments on study results, it would thus be desirable to conduct experiments with and without these adjustments in the same participants in two separate sessions (although this necessitates control of possible learning effects between the two test sessions).

Third, the present thesis tested participant's neural and behavioural dynamics of attention mechanisms exclusively in auditory (speech) paradigms. It is thus an open question how specific the observed results are to the auditory modality. For instance, it might be that similar alpha power modulations as in Study 2.1 would also be observed for the degradation of visual or somatosensory stimuli. Since alpha power modulations have been observed across modalities (somatosensation: van Ede et al., 2011; audition: Weisz et al., 2011; vision: Bauer et al., 2012), it is likely that they constitute a general neural mechanism to control attention in different sensory modalities. A thorough differentiation of neural signatures in the M/EEG between different modalities would benefit from the combination with neuroimaging methods with a higher spatial resolution such as fMRI (e.g., Debener et al., 2006; Walz et al., 2013; Scharinger et al., 2014). Thus, combined M/EEG-fMRI studies could reveal whether neural signatures of attention are functionally similar across modalities, but localized to sensory-specific brain regions depending on the stimulus modality.

7.6 A framework for the role of attention for speech comprehension under demanding acoustic conditions

Put simply, acoustic degradation decreases speech comprehension, whereas enhanced attention might improve speech comprehension. In the present thesis, we found that neural signatures in the M/EEG reflect demands on attention, as well as the use of acoustic and predictive cues to overcome acoustic degradation. Importantly, modulation of neural activity also predicts the success of speech comprehension in demanding listening tasks (Studies 1.1, 2.1, & 3). In this section, I will argue that acoustic degradation and attention not only have opposing effects speech comprehension, but that enhanced attention can compensate for the detrimental effects of acoustic degradation.

Figure 7.1 illustrates the proposed framework for the role of attention for speech comprehension under acoustic degradation. The major presumption of this framework is that speech comprehension depends on the ratio between acoustic degradation and attentional control (x-axis in Figure 7.1B: attentional control/acoustic degradation). Attentional control describes the volitional (i.e., top-down) selection of a task-relevant signal and the suppression of task-irrelevant noise (e.g., Hill and Miller, 2010). Attentional control can vary both on a broader time scale as a function of age (e.g., Chao and Knight, 1997; Passow et al., 2012) but also from one moment to the next due to e.g., lapsing attention (O'Connell et al., 2009). Acoustic degradation in the present thesis refers to the deterioration of an acoustic input signal through masking with noise (e.g., Brungart et al., 2001) or through the distortion of spectral features (e.g., Shannon et al., 1995). The proposed framework applies to listening situations involving some kind of acoustic degradation but not directly to ideal listening conditions in a silent acoustic background.

In general, three areas concerning the ratio of attentional control and acoustic degradation can be differentiated. First, if acoustic degradation dominates over attentional control, speech comprehension is poor (red area in Figure 7.1). Second, speech comprehension is most dynamic in contexts where acoustic degradation and attentional control counterbalance each other (yellow area in Figure 7.1). Third, if attentional control dominates over acoustic degradation, speech comprehension is effortless and attention resources are eventually available for other tasks (green area in Figure 7.1). In the following, I will outline in how far experimental results from the present thesis support this framework.

Figure 7.1B assigns results from this thesis to the three areas of the proposed framework. It is obvious that most of our research targeted the area where speech comprehension is most dynamic, i.e., the area where the negative effects of acoustic degradation and the compensatory mechanisms of attentional control are relatively balanced (yellow area in Figure 7.1). This has mainly practical reasons since experimental manipulations in this area most likely lead to observable effects on speech comprehension. But to what extent do our results support the opposing effects of acoustic degradation and attentional control in this area? Three studies of this thesis found evidence that stronger modulations of neural attention mechanisms correlate with improved speech comprehension. First, larger magnitude of the contingent negative variation (CNV) predicted better performance in the auditory number comparison task (see Figure 3.4). Second, listeners with stronger alpha power modulations as a function of acoustics and predictiveness of speech reported lower effort of speech comprehension in demanding listening situations in everyday life (see Figure 4.5A). Third, listeners who showed stronger modulations of alpha power lateralization at the presentation rate of two concurrent speech streams made fewer errors in the report of task-relevant speech in a dichotic listening task (see Figure 5.5D). Taken together, these findings demonstrate that neural attention mechanisms can counteract acoustic degradation to support speech comprehension.

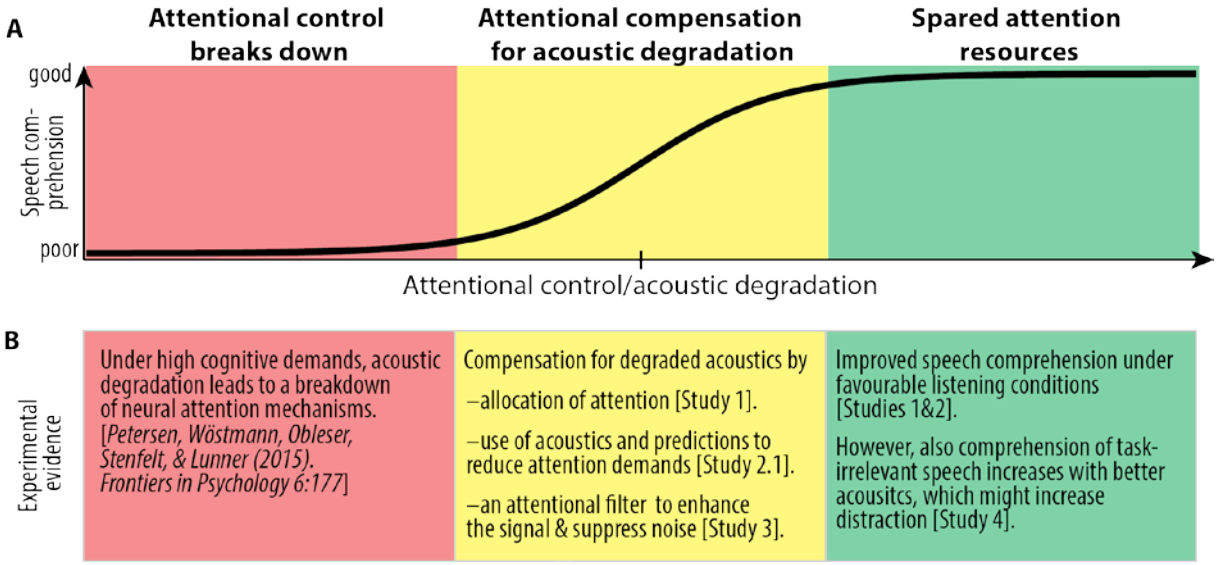

Figure 7.1. Framework of the role of attention for speech comprehension under demanding acoustic conditions.(A) An initial framework explaining how the ratio of attentional control and acoustic degradation shapes speech comprehension. See text for a thorough explanation of this framework. (B) Assignment of experimental results from the present thesis to the three areas of the framework.

However, what happens to the neural dynamics of attention in the case of a significant disbalance between acoustic degradation and attentional control (i.e., in the red and green areas in Figure 7.1)? In a recent collaborative study with researchers at the Eriksholm research centre in Denmark, we found

initial evidence for a breakdown of neural attention mechanisms if acoustic degradation exceeds a certain "breakpoint" (Petersen et al., 2015). In detail, older hearing-impaired listeners had to remember a variable number of spoken digits under varying levels of background noise. In the highest memory load condition (i.e., six digits to be remembered), alpha power increased with the severity of hearing loss for lower levels of background noise. However, in the strongest background noise condition, alpha power increased only up to intermediate levels of hearing loss but then decreased for listeners with the most severe hearing loss. This result might indicate that neural mechanisms of attention compensate for acoustic degradation up to a certain point (i.e., the "breakpoint"). But if acoustic degradation becomes too strong, neural mechanisms of attention break down (see also Reuter-Lorenz and Cappell, 2008). It is important to note that evidence for this breakdown of attentional mechanisms for speech comprehension is limited at present. Future studies should further investigate the relationship between the breakdown of neural attention mechanisms and behavioural measures of speech comprehension.

What are the dynamics of speech comprehension in a favourable listening situation, where attentional control dominates over acoustic degradation (green are in Figure 7.1)? Studies 1&2 revealed that under most favourable listening conditions (high level of preserved temporal fine structure), objective measures (% correct and RTs) and subjective measures of performance (confidence ratings) indicated improved speech comprehension. Thus, speech comprehension benefits from a higher ratio between attentional control and acoustic degradation. However, more preserved acoustic detail also enhances the attentional capture of task-irrelevant speech which might compromise attention to task-relevant speech (Study 4). This is in agreement with the theory that under reduced perceptual load (e.g., less degraded acoustics), processing of task-irrelevant distractors increases (Lavie et al., 2004; Lavie, 2005). In general, there is at present a lack of research on the mechanisms of attention for speech comprehension in more favourable listening conditions where attentional control outweighs acoustic degradation (green area in Figure 7.1). One possibility is that attention resources otherwise deployed to support speech comprehension become available for other tasks when the ratio between attentional control and acoustic degradation increases (for a simillar theory on working memory resources, see Mishra et al., 2014; Rudner and Lunner, 2014).

How does the proposed framework account for the increase in listening difficulties for older adults with progressive hearing loss (Pichora-Fuller and Souza, 2003)? Age-related hearing loss increases degradation of the acoustic input which in turn enhances the need for attentional control (Shinn-

Cunningham and Best, 2008). Thus, older listeners have to upregulate attention mechanisms more than younger listeners to ensure speech comprehension(see also Erb and Obleser, 2013). However, since attentional control declines at an older age (Chao and Knight, 1997; Passow et al., 2012), compensation for acoustic degradation might fail. Thus, both hearing loss and cognitive decline foster the dominance of acoustic degradation over attentional control, causing decreased speech comprehension in the elderly. Our framework suggests a general tendency that older listeners with impaired hearing end up at lower ratios between attentional control and acoustic degradation (i.e., in the red are in Figure 7.1).

Finally, I would like to outline predictions arising from the proposed framework which could be tested in future studies. I will here explain three testable hypotheses, one for each of the three areas in the framework. First, our framework asserts that if acoustic degradation outweighs the compensatory function of attentional control, neural mechanisms of attention break down (red area in Figure 7.1). Future studies could investigate in how far the breakpoint of neural attention mechanisms depends on acoustic conditions and individual capabilities of attentional control. In an attention demanding listening task with varying levels of task difficulty, our framework hypothesizes that acoustic degradation moves the breakpoint to lower levels of task-difficulty. In contrast, the breakpoint should move to higher levels of task difficulty for listeners with better attentional control (quantifiable through neuropsychological markers such as the d2-R test, see section 3.1.2.5).

Second, the present thesis provides good evidence that stronger modulations of neural attention mechanisms (e.g., CNV and alpha power) support speech comprehension and compensate for degraded acoustics (yellow area in Figure 7.1). Critically, our framework implies that this principle does not only hold in laboratory settings but also in everyday life communication situations which are arguably more dynamic. This could be tested with a mobile EEG system, which assesses well-defined neural signatures while the listener is engaged in an effortful listening situation. For instance, EEG measures could assess listeners' current attentional effort (reflected by e.g., CNV and alpha power) as well as listeners' current focus of attention in a multi-talker environment (reflected by e.g., alpha power lateralization). Our framework supposes that an active manipulation of the acoustic conditions would affect neural signatures of attention and also speech comprehension. In detail, acoustic conditions could be manipulated in a way that neural signatures indicate decreasing demands on attention. This should be accompanied by improved speech comprehension. For instance, if lateralized alpha power indicates a rightward shift of the listener' focus of attention, amplification of the acoustic

input to the right ear should reduce demands on attention and improve speech comprehension. This could be realized through bilateral hearing aids which are connected to the EEG system. Future research could thus implement such a "closed loop" between EEG recordings and adjustable hearing aids. In general, our framework predicts that whenever adjustments of hearing aids cause a decrease in neural signatures of demands on attention, speech comprehension should improve.

Third, if high attentional control (e.g., in a younger healthy listeners) dominates over a mild acoustic degradation, our framework predicts that attention capacities are not entirely exploited by the listening task but can be used for other tasks (green area in Figure 7.1). This hypothesis could be investigated in a between-subject study using a dual-task paradigm. It has been shown that individuals differ in the extent to which speech comprehension in noise benefits from temporal fine structure (e.g., Hopkins et al., 2008; Lunner et al., 2012; Neher et al., 2012). One possible interpretation might be that listeners who profit strongly from temporal fine structure consume fewer attention capacities for speech comprehension if fine structure is preserved (for a nuanced discussion of the role of temporal fine structure for speech comprehension, see Study 1.1). According to our framework, these spare attention capacities could potentially improve performance on concurrent attention-demanding tasks. To test this, participants could perform a speech comprehension task under varying levels of temporal fine structure. Concurrently, they would have the task to count randomly presented brief tones. Our framework hypothesizes a positive relationship between participants' benefit from temporal fine structure in the speech comprehension task and accuracy in the tone counting task. That is, listeners who benefit a lot from fine structure in the speech comprehension task free more attention capacities and thus improve performance also in the concurrent tone counting task. In contrast, listeners who do not profit from more preserved fine structure cannot free attention capacity and exhibit no performance increase in the tone counting task. This would demonstrate that attention capacities spared under a high ratio between attentional control and acoustic degradation can be exploited for other attention-demanding tasks.

Taken together, the framework proposed here accounts for the results of this thesis and generates testable hypotheses for future studies. In order to test this framework and to foster our understanding of how neural mechanisms shape speech comprehension, I consider it inevitable to directly investigate the relationship between neural mechanisms and behavioural measures of speech comprehension. Unfortunately, neural dynamics and behavioural outcomes are often analysed in isolation or in parallel without using the great potential of relating these two to one another. Thus, I would finally like to

make the point here that a thorough understanding of how neural dynamics support speech comprehension can best be achieved by relating these neural dynamics to listeners' behavioural performance of speech comprehension.

8 Conclusions

This thesis used behavioural and electrophysiological measures to investigate the neural dynamics of selective attention to speech under demanding acoustic conditions, such as spectral degradation and the presence of background noise. In the following, I will conclude that findings of this thesis (1) foster our understanding of how the neural system implements selective attention to speech in noise, (2) provide insights in the neural bases of inter-individual differences in listening difficulties, particularly at an older age, and (3) set the stage for future interventions to support speech comprehension under attention-demanding acoustic conditions.

Results of this thesis indicate that selective attention to speech in noise is not implemented by a single but rather by several neural mechanisms that complement each other. Slow cortical potentials reflect the anticipatory allocation of attention under acoustic degradation; neural phase-locking to the acoustic envelope segregates attended and unattended speech on a neural level; and alpha oscillations indicate the inhibition of irrelevant brain regions and processes to support processing to task-relevant speech. Our results thus emphasizes that humans' remarkable ability to volitionally select one speech stream from the plethora of sounds arriving at the ear in a complex listening situation depends on temporally and functionally distinct neural mechanisms. Since these neural mechanisms not only reflect demands on attention but also explain listeners' success of attentional selection, they provide a functionally significant neural basis of speech comprehension under demanding acoustic conditions.

This thesis uncovers in how far changes in neural dynamics might explain the listening difficulties experienced at an older age. Our findings demonstrate that the anticipatory allocation of attention and the differential neural phase-locking to attended and unattended speech are preserved at an older age, given that acoustic conditions are carefully adapted to individual requirements. This emphasizes the need for dynamic and individualized approaches to adapt acoustic conditions to hearing acuity and cognitive functioning. However, even with these individual adjustments, older compared with younger listeners are stronger driven by bottom-up acoustic features of the external stimulation. This might explain why older listeners experience particularly strong interference from task-irrelevant background noise in multi-talker situations.

Finally, results of this thesis provide a basis for an initial framework which holds that neural mechanisms of attention support speech comprehension and thus counteract the detrimental effects of

acoustic degradation (for details, see section 7.6). Critically, the interdependence between the degradation of the acoustic input and neural mechanisms of attention opens up the possibility to understand how the acoustic input could be manipulated to reduce listeners' demands on attention. In detail, manipulations of sound processing procedures in hearing aids could be correlated with neural signatures of attention to find those hearing aid parameters that minimize demands on attention and thus improve speech comprehension. It is therefore a timely endeavour to utilize neural signatures of attention to speech in noise for interventions that aim at improving the individual listener's abilities of speech comprehension under demanding acoustic conditions.

References

Adrian ED (1944) Brain Rhythms. Nature 153:360-362.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367-13372.

Ahlfors SP, Han J, Belliveau JW, Hamalainen MS (2010) Sensitivity of MEG and EEG to source orientation. Brain Topogr 23:227-232.

Ahveninen J, Huang S, Belliveau JW, Chang WT, Hamalainen M (2013) Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. J Cogn Neurosci 25:1926-1943.

Alho K, Salonen J, Rinne T, Medvedev SV, Hugdahl K, Hamalainen H (2012) Attention-related modulation of auditory-cortex responses to speech sounds during dichotic listening. Brain Res 1442:47-54.

Arbogast TL, Mason CR, Kidd G, Jr. (2005) The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 117:2169-2180.

Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390-398.

Attneave F, Olson RK (1971) Pitch as a medium: a new approach to psychophysical scaling. Am J Psychol 84:147-166.

Awh E, Vogel EK, Oh SH (2006) Interactions between attention and working memory. Neuroscience 139:201-208.

Baddeley A (1992) Working memory. Science 255:556-559.

Baddeley A, Hitch G (1974) Working Memory. In: The psychology of learning and motivation (Bower GH, ed), pp 47-89. New York, London: Academic Press.

Baddeley A, Salame P (1986) The unattended speech effect: perception or memory? J Exp Psychol Learn Mem Cogn 12:525-529.

Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. Science 344:424-427.

Baltes PB, Lindenberger U (1997) Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? Psychol Aging 12:12-21.

Banerjee S, Snyder AC, Molholm S, Foxe JJ (2011) Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms? J Neurosci 31:9923-9932.

Bates ME, Lemay EP, Jr. (2004) The d2 Test of attention: construct validity and extensions in scoring techniques. J Int Neuropsychol Soc 10:392-400.

Bauer M, Kennett S, Driver J (2012) Attentional selection of location and modality in vision and touch modulates low-frequency activity in associated sensory cortices. J Neurophysiol 107:2342-2351.

Becker R, Pefkou M, Michel CM, Hervais-Adelman AG (2013) Left temporal alpha-band activity reflects single word intelligibility. Front Syst Neurosci 7:121.

Bell R, Buchner A, Mund I (2008) Age-related differences in irrelevant-speech effects. Psychol Aging 23:377-391.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological) 57:289-300.

Berger H (1931) Über das Elektrenkephalogramm des Menschen. Archiev für Psychiatrie und Nervenkrankheiten 94:16-60.

Best V, Ozmeral EJ, Kopco N, Shinn-Cunningham BG (2008) Object continuity enhances selective auditory attention. Proc Natl Acad Sci U S A 105:13174-13178.

Bhatt KA, Liberman MC, Nadol JB, Jr. (2001) Morphometric analysis of age-related changes in the human basilar membrane. Ann Otol Rhinol Laryngol 110:1147-1153.

Bonnefond M, Jensen O (2012) Alpha oscillations serve to protect working memory maintenance against anticipated distracters. Curr Biol 22:1969-1974.

Bosman CA, Schoffelen JM, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P (2012) Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron 75:875-888.

Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci 8:539-546.

Brant LJ, Fozard JL (1990) Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. J Acoust Soc Am 88:813-820.

Brefczynski JA, DeYoe EA (1999) A physiological correlate of the 'spotlight' of visual attention. Nat Neurosci 2:370-374.

Brickenkamp R, Schmidt-Atzert L, Liepmann D (2010) d2–Revision Aufmerksamkeits-und Konzentrationstest. In. Göttingen, Hogrefe Verlag.

Briggs F, Mangun GR, Usrey WM (2013) Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature 499:476-480.

Broadbent D (1958) Perception and Communication. London: Pergamon.

Broyd SJ, Richards HJ, Helps SK, Chronaki G, Bamford S, Sonuga-Barke EJ (2012) An electrophysiological monetary incentive delay (e-MID) task: a way to decompose the different components of neural response to positive and negative monetary reinforcement. J Neurosci Methods 209:40-49.

Brungart DS, Simpson BD, Ericson MA, Scott KR (2001) Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am 110:2527-2538.

BSA (2011) Recommended Procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. In. UK: Berkshire.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. Proc Natl Acad Sci U S A 108:11262-11267.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186-198.

Busch NA, Dubois J, VanRullen R (2009) The phase of ongoing EEG oscillations predicts visual perception. J Neurosci 29:7869-7876.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A 102:18751-18756.

Buzsaki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926-1929.

Buzsaki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407-420.

Carlson MC, Hasher L, Zacks RT, Connelly SL (1995) Aging, distraction, and the benefits of predictable location. Psychol Aging 10:427-436.

Cartwright-Finch U, Lavie N (2007) The role of perceptual load in inattentional blindness. Cognition 102:321-340.

Chao LL, Knight RT (1997) Prefrontal deficits in attention and inhibitory control with aging. Cereb Cortex 7:63-69.

Chen Z (2012) Object-based attention: a tutorial review. Atten Percept Psychophys 74:784-802.

Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibanez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. J Neurosci 33:11194-11205.

Cherry CE (1953) Some Experiments on the Recognition of Speech, with One and with Two Ears. Journal of the Acoustical Society of America 25:975-979.

Clinard CG, Tremblay KL, Krishnan AR (2010) Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. Hear Res 264:48-55.

Cohen D (1968) Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. Science 161:784-786.

Cohen J (1969) Statistical power analysis for the behavioural sciences. New York: Academic Press.

Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12:1594-1600.

Colflesh GJ, Conway AR (2007) Individual differences in working memory capacity and divided attention in dichotic listening. Psychon Bull Rev 14:699-703.

Colle HA, Welsh A (1976) Acoustic Masking in Primary Memory. Journal of Verbal Learning and Verbal Behavior 15:17-31.

Conway AR, Cowan N, Bunting MF (2001) The cocktail party phenomenon revisited: the importance of working memory capacity. Psychon Bull Rev 8:331-335.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201-215.

Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, Shulman GL (1998) A common network of functional areas for attention and eye movements. Neuron 21:761-773.

Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci 24:87-114.

Cravo AM, Rohenkohl G, Wyart V, Nobre AC (2013) Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. J Neurosci 33:4002-4010.

Damasio AR, Damasio H, Chui HC (1980) Neglect following damage to frontal lobe or basal ganglia. Neuropsychologia 18:123-132.

Dau T, Kollmeier B, Kohlrausch A (1997) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J Acoust Soc Am 102:2892-2905.

Debener S, Ullsperger M, Siegel M, Engel AK (2006) Single-trial EEG-fMRI reveals the dynamics of cognitive function. Trends Cogn Sci 10:558-563.

Dehaene S, Dehaene-Lambertz G, Cohen L (1998) Abstract representations of numbers in the animal and human brain. Trends Neurosci 21:355-361.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193-222.

Deutsch JA, Deutsch D (1963) Attention: Some theoretical considerations. Psychol Rev 70:80-90.

Dienes Z (2014) Using Bayes to get the most out of non-significant results. Front Psychol 5:781.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109:11854-11859.

Ding N, Chatterjee M, Simon JZ (2013) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88C:41-46.

Doppelmayr MM, Klimesch W, Pachinger T, Ripper B (1998) The functional significance of absolute power with respect to event-related desynchronization. Brain Topogr 11:133-140.

Driver J, Mattingley JB (1998) Parietal neglect and visual awareness. Nat Neurosci 1:17-22.

Dugue L, Marque P, VanRullen R (2011) The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. J Neurosci 31:11889-11893.

Duncan J (1984) Selective attention and the organization of visual information. J Exp Psychol Gen 113:501-517.

Durlach NI, Mason CR, Kidd G, Jr., Arbogast TL, Colburn HS, Shinn-Cunningham BG (2003) Note on informational masking. J Acoust Soc Am 113:2984-2987.

Edward DL, Ryan MR, Koestner R (1999) A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. Psychological Bulletin 125:627-668.

Ellermeier W, Kattner F, Ueda K, Doumoto K, Nakajima Y (2012) Effects of noise-vocoded speech in the irrelevant-sound paradigm. In: 28. Jahrestagung für Akustik, pp 336-337. Berlin: Galrev.

Engelmann JB, Pessoa L (2007) Motivation sharpens exogenous spatial attention. Emotion 7:668-674.

Erb J, Obleser J (2013) Upregulation of cognitive control networks in older adults' speech comprehension. Front Syst Neurosci 7:116.

Erb J, Henry MJ, Eisner F, Obleser J (2012) Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. Neuropsychologia 50:2154-2164.

Faulkner A, Rosen S, Wilkinson L (2001) Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. Ear Hear 22:431-438.

Fischl B, Sereno MI, Dale AM (1999a) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195-207.

Fischl B, Sereno MI, Tootell RB, Dale AM (1999b) High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 8:272-284.

Fisk JE, Warr P (1996) Age and working memory: the role of perceptual speed, the central executive, and the phonological loop. Psychol Aging 11:316-323.

Fostick L, Babkoff H (2013) Temporal and non-temporal processes in the elderly. J Basic Clin Physiol Pharmacol 24:191-199.

Foxe JJ, Snyder AC (2011) The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. Front Psychol 2:154.

Foxe JJ, Simpson GV, Ahlfors SP (1998) Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. Neuroreport 9:3929-3933.

Frey JN, Mainy N, Lachaux JP, Muller N, Bertrand O, Weisz N (2014) Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. J Neurosci 34:6634-6639.

Frisina RD (2009) Age-related hearing loss: ear and brain mechanisms. Ann N Y Acad Sci 1170:708-717.

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention--focusing the searchlight on sound. Curr Opin Neurobiol 17:437-455.

Fu KM, Foxe JJ, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2001) Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto-occipital alpha-band oscillations. Brain Res Cogn Brain Res 12:145-152.

Füllgrabe C (2013) Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. Am J Audiol 22:313-315.

Furman AC, Kujawa SG, Liberman MC (2013) Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. J Neurophysiol 110:577-586.

Gandhi SP, Heeger DJ, Boynton GM (1999) Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci U S A 96:3314-3319.

Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2011) The importance of being variable. J Neurosci 31:4496-4503.

Gazzaley A, Cooney JW, Rissman J, D'Esposito M (2005a) Top-down suppression deficit underlies working memory impairment in normal aging. Nat Neurosci 8:1298-1300.

Gazzaley A, Sheridan MA, Cooney JW, D'Esposito M (2007) Age-related deficits in component processes of working memory. Neuropsychology 21:532-539.

Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M (2005b) Top-down enhancement and suppression of the magnitude and speed of neural activity. J Cogn Neurosci 17:507-517.

Gazzaley A, Clapp W, Kelley J, McEvoy K, Knight RT, D'Esposito M (2008) Age-related top-down suppression deficit in the early stages of cortical visual memory processing. Proc Natl Acad Sci U S A 105:13122-13126.

Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511-517.

Gisselgard J, Petersson KM, Ingvar M (2004) The irrelevant speech effect and working memory load. Neuroimage 22:1107-1116.

Goldstein RZ, Cottone LA, Jia Z, Maloney T, Volkow ND, Squires NK (2006) The effect of graded monetary reward on cognitive event-related potentials and behavior in young healthy adults. Int J Psychophysiol 62:272-279.

Gomez-Ramirez M, Kelly SP, Molholm S, Sehatpour P, Schwartz TH, Foxe JJ (2011) Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. J Neurosci 31:18556-18567.

Grose JH, Mamo SK (2010) Processing of temporal fine structure as a function of age. Ear Hear 31:755-760.

Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci U S A 98:694-699.

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6:e138.

Haegens S, Handel BF, Jensen O (2011a) Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. J Neurosci 31:5197-5204.

Haegens S, Nacher V, Luna R, Romo R, Jensen O (2011b) alpha-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proc Natl Acad Sci U S A 108:19377-19382.

Hämäläinen MS (1995) Functional localization based on measurements with a whole-head magnetometer system. Brain Topogr 7:283-289.

Hambrook DA, Tata MS (2014) Theta-band phase tracking in the two-talker problem. Brain Lang 135:52-56.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32:180-194.

Hanslmayr S, Aslan A, Staudigl T, Klimesch W, Herrmann CS, Bauml KH (2007) Prestimulus oscillations predict visual perception performance between and within subjects. Neuroimage 37:1465-1473.

Hartwigsen G, Golombek T, Obleser J (2014) Repetitive transcranial magnetic stimulation over left angular gyrus modulates the predictability gain in degraded speech comprehension. Cortex, doi: 10.1016/j.cortex.2014.08.027.

He BJ, Raichle ME (2009) The fMRI signal, slow cortical potential and consciousness. Trends Cogn Sci 13:302-309.

Heinze HJ, Luck SJ, Mangun GR, Hillyard SA (1990) Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. Electroencephalogr Clin Neurophysiol 75:511-527.

Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proc Natl Acad Sci U S A 109:20095-20100.

Herrmann B, Henry MJ, Scharinger M, Obleser J (2014) Supplementary motor area activations predict individual differences in temporal-change sensitivity and its illusory distortions. Neuroimage, doi:10.1016/j.neuroimage.2014.07.026.

Herrmann CS, Knight RT (2001) Mechanisms of human attention: event-related potentials and oscillations. Neurosci Biobehav Rev 25:465-476.

Hertrich I, Dietrich S, Trouvain J, Moos A, Ackermann H (2012) Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49:322-334.

Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cereb Cortex 20:583-590.

Hillebrand A, Barnes GR (2002) A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. Neuroimage 16:638-650.

Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the human brain. Science 182:177-180.

Hinterberger T, Veit R, Wilhelm B, Weiskopf N, Vatine JJ, Birbaumer N (2005) Neuronal mechanisms underlying control of a brain-computer interface. Eur J Neurosci 21:3169-3181.

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3:284-291.

Hopkins K, Moore BC (2009) The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. J Acoust Soc Am 125:442-446.

Hopkins K, Moore BC (2010) The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects. J Acoust Soc Am 127:1595-1608.

Hopkins K, Moore BC (2011) The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. J Acoust Soc Am 130:334-349.

Hopkins K, Moore BC, Stone MA (2008) Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. J Acoust Soc Am 123:1140-1153.

Horton C, D'Zmura M, Srinivasan R (2013) Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol 109:3082-3093.

Huang-Pollock CL, Carr TH, Nigg JT (2002) Development of selective attention: perceptual load influences early versus late attentional selection in children and adults. Dev Psychol 38:363-375.

Hugdahl K, Westerhausen R, Alho K, Medvedev S, Laine M, Hamalainen H (2009) Attention and cognitive control: unfolding the dichotic listening story. Scand J Psychol 50:11-22.

Humes LE (1996) Speech understanding in the elderly. J Am Acad Audiol 7:161-167.

Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res 40:1489-1506.

Jackson AF, Bolger DJ (2014) The neurophysiological bases of EEG and EEG measurement: a review for the rest of us. Psychophysiology 51:1061-1071.

Jennings CR, Jones NS (2001) Presbyacusis. J Laryngol Otol 115:171-178.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn Sci 16:200-206.

Jensen O, Gelfand J, Kounios J, Lisman JE (2002) Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. Cereb Cortex 12:877-882.

Jepsen ML, Ewert SD, Dau T (2008) A computational model of human auditory signal processing and perception. J Acoust Soc Am 124:422-438.

Johnson JA, Zatorre RJ (2005) Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. Cereb Cortex 15:1609-1620.

Jones D, Morris N (1992) Irrelevant speech and serial recall: implications for theories of attention and working memory. Scand J Psychol 33:212-229.

Jones DM, Macken WJ (1993) Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. Journal of Experimental Psychology 19:369-381.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541-577.

Kalikow DN, Stevens KN, Elliott LL (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 61:1337-1351.

Karnath HO, Ferber S, Himmelbach M (2001) Spatial awareness is a function of the temporal not the posterior parietal lobe. Nature 411:950-953.

Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282:108-111.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751-761.

Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party". J Neurosci 30:620-628.

Kidd G, Jr., Arbogast TL, Mason CR, Gallun FJ (2005) The advantage of knowing where to listen. J Acoust Soc Am 118:3804-3815.

Kim YH, Gitelman DR, Nobre AC, Parrish TB, LaBar KS, Mesulam MM (1999) The large-scale neural network for spatial attention displays multifunctional overlap but differential asymmetry. Neuroimage 9:269-277.

Kitayama S (1991) Impairment of perception by positive and negative affect. Cognition & Emotion 5:255-274.

Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Brain Res Rev 29:169-195.

Klimesch W (2012) alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn Sci 16:606-617.

Klimesch W, Sauseng P, Gerloff C (2003) Enhancing cognitive performance with repetitive transcranial magnetic stimulation at human individual alpha frequency. Eur J Neurosci 17:1129-1133.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev 53:63-88.

Klimesch W, Doppelmayr M, Pachinger T, Ripper B (1997) Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. Neurosci Lett 238:9-12.

Klimesch W, Doppelmayr M, Russegger H, Pachinger T, Schwaiger J (1998) Induced alpha band power changes in the human EEG and attention. Neurosci Lett 244:73-76.

Klimesch W, Doppelmayr M, Schwaiger J, Auinger P, Winkler T (1999) 'Paradoxical' alpha synchronization in a memory task. Brain Res Cogn Brain Res 7:493-501.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. J Neurosci 25:4806-4812.

Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197-2206.

Kong YY, Mullangi A, Ding N (2014) Differential modulation of auditory responses to attended and unattended speech in different listening conditions. Hear Res 316:73-81.

Kopp F, Schröger E, Lipka S (2004) Neural networks engaged in short-term memory rehearsal are disrupted by irrelevant speech in human subjects. Neurosci Lett 354:42-45.

Kruschke JK (2011) Bayesian assessment of null values via parameter estimation and model comparison. Perspectives on Psychological Sciences 6:299-312.

Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80:97-126.

Kujawa SG, Liberman MC (2009) Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J Neurosci 29:14077-14085.

Kutas M, Hillyard SA (1980) Event-related brain potentials to semantically inappropriate and surprisingly large words. Biol Psychol 11:99-116.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194-208.

Lachter J, Forster KI, Ruthruff E (2004) Forty-five years after Broadbent (1958): still no identification without attention. Psychol Rev 111:880-913.

Lakatos P, Schroeder CE, Leitman DI, Javitt DC (2013a) Predictive suppression of cortical excitability and its deficit in schizophrenia. J Neurosci 33:11692-11702.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904-1911.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013b) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750-761.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799-1822.

Large EW, Jones MR (1999) The dynamics of attending: How people track time-varying events. Psychological Review 106:119-159.

Laufs H, Kleinschmidt A, Beyerle A, Eger E, Salek-Haddadi A, Preibisch C, Krakow K (2003) EEG-correlated fMRI of human alpha activity. Neuroimage 19:1463-1476.

Lavie N (2005) Distracted and confused?: selective attention under load. Trends Cogn Sci 9:75-82.

Lavie N (2010) Attention, Distraction, and Cognitive Control Under Load. Current Directions in Psychological Science 19:143-148.

Lavie N, Hirst A, de Fockert JW, Viding E (2004) Load theory of selective attention and cognitive control. J Exp Psychol Gen 133:339-354.

Lawrence NS, Ross TJ, Hoffmann R, Garavan H, Stein EA (2003) Multiple neuronal networks mediate sustained attention. J Cogn Neurosci 15:1028-1038.

Leahy RM, Mosher JC, Spencer ME, Huang MX, Lewine JD (1998) A study of dipole localization accuracy for MEG and EEG using a human skull phantom. Electroencephalogr Clin Neurophysiol 107:159-173.

LeCompte DC, Neely CB, Wilson JR (1997) Irrelevant speech and irrelevant tones: the relative importance of speech to the irrelevant speech effect. J Exp Psychol Learn Mem Cogn 23:472-483.

Lehtela L, Salmelin R, Hari R (1997) Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex. Neurosci Lett 222:111-114.

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:Suppl 2:467+.

Li KZ, Lindenberger U (2002) Relations between aging sensory/sensorimotor and cognitive functions. Neurosci Biobehav Rev 26:777-783.

Lopes da Silva F (2013) EEG and MEG: relevance to neuroscience. Neuron 80:1112-1128.

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A 103:18866-18869.

Loveless NE, Sanford AJ (1974) Effects of age on the contingent negative variation and preparatory set in a reaction-time task. J Gerontol 29:52-63.

Luck SJ, Woodman GF, Vogel EK (2000) Event-related potential studies of attention. Trends Cogn Sci 4:432-440.

Luck SJ, Heinze HJ, Mangun GR, Hillyard SA (1990) Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. Electroencephalogr Clin Neurophysiol 75:528-542.

Lunner T, Rudner M, Ronnberg J (2009) Cognition and hearing aids. Scand J Psychol 50:395-403.

Lunner T, Hietkamp RK, Andersen MR, Hopkins K, Moore BC (2012) Effect of speech material on the benefit of temporal fine structure information in speech for young normal-hearing and older hearing-impaired participants. Ear Hear 33:377-388.

Malmivuo J (2012) Comparison of the properties of EEG and MEG in detecting the electric activity of the brain. Brain Topogr 25:1-19.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177-190.

Marois R, Ivanoff J (2005) Capacity limits of information processing in the brain. Trends Cogn Sci 9:296-305.

Mazaheri A, van Schouwenburg MR, Dimitrijevic A, Denys D, Cools R, Jensen O (2014) Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. Neuroimage 87:356-362.

McCallum WC, Walter WG (1968) The effects of attention and distraction on the contingent negative variation in normal and neurotic subjects. Electroencephalogr Clin Neurophysiol 25:319-329.

McCoy SL, Tun PA, Cox LC, Colangelo M, Stewart RA, Wingfield A (2005) Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. Q J Exp Psychol A 58:22-33.

McDermott JH, Schemitsch M, Simoncelli EP (2013) Summary statistics in auditory perception. Nat Neurosci 16:493-498.

Meister H, Schreitmuller S, Grugel L, Ortmann M, Beutner D, Walger M, Meister IG (2012) Cognitive resources related to speech recognition with a competing talker in young and older listeners. Neuroscience 232C:74-82.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233-236.

Meyer L, Obleser J, Friederici AD (2013) Left parietal alpha enhancement during working memory-intensive sentence processing. Cortex 49:711-721.

Miller GA (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81-97.

Mishra S, Stenfelt S, Lunner T, Ronnberg J, Rudner M (2014) Cognitive spare capacity in older adults with hearing loss. Front Aging Neurosci 6:96.

Moore BC (2008a) Basic auditory processes involved in the analysis of speech sounds. Philos Trans R Soc Lond B Biol Sci 363:947-963.

Moore BC (2008b) The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J Assoc Res Otolaryngol 9:399-406.

Moore BC, Alcantara JI, Glasberg BR (1998) Development and evaluation of a procedure for fitting multi-channel compression hearing aids. Br J Audiol 32:177-195.

Moore BC, Glasberg BR, Stoev M, Fullgrabe C, Hopkins K (2012) The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L). J Acoust Soc Am 131:1003-1006.

Moray N (1959) Attention in dichotic listening: Affective cues and the influence of instructions. Quarterly Journal of Experimental Psychology 11:56-60.

Morton J, Macrus S, Frankish C (1976) Perceptual centers (P-centers). Psychological Review 83:405-408.

Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol 70:909-919.

Moyer RS, Landauer TK (1967) Time required for judgements of numerical inequality. Nature 215:1519-1520.

Müller N, Weisz N (2012) Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. Cereb Cortex 22:1604-1613.

Murphy DR, McDowd JM, Wilcox KA (1999) Inhibition and aging: similarities between younger and older adults as revealed by the processing of unattended auditory information. Psychol Aging 14:44-59.

Näätänen R, Gaillard AW, Mantysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst) 42:313-329.

Näätänen R, Gaillard AW, Varey CA (1981) Attention effects on auditory EPs as a function of inter-stimulus interval. Biol Psychol 13:173-187.

Nagai Y, Critchley HD, Featherstone E, Fenwick PB, Trimble MR, Dolan RJ (2004) Brain activity relating to the contingent negative variation: an fMRI investigation. Neuroimage 21:1232-1241.

Neher T, Lunner T, Hopkins K, Moore BC (2012) Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L). J Acoust Soc Am 131:2561-2564.

Neher T, Behrens T, Carlile S, Jin C, Kragelund L, Petersen AS, Schaik A (2009) Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition. Int J Audiol 48:758-774.

Nilsson LG (2003) Memory function in normal aging. Acta Neurol Scand Suppl 179:7-13.

Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol 48:3637-3652.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564-15574.

O'Connell RG, Dockree PM, Robertson IH, Bellgrove MA, Foxe JJ, Kelly SP (2009) Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur. J Neurosci 29:8604-8611.

O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5:1203-1209.

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC (2014) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex.

Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb Cortex 22:2466-2477.

Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci 28:8116-8123.

Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283-2289.

Obleser J, Wöstmann M, Hellbernd N, Wilsch A, Maess B (2012) Adverse listening conditions and memory load drive a common alpha oscillatory network. J Neurosci 32:12376-12383.

Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C (2007) Attention improves population-level frequency tuning in human auditory cortex. J Neurosci 27:10383-10390.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Osipova D, Hermes D, Jensen O (2008) Gamma power is phase-locked to posterior alpha activity. PLoS One 3:e3990.

Palva JM, Palva S, Kaila K (2005) Phase synchrony among neuronal oscillations in the human cortex. J Neurosci 25:3962-3972.

Park HL, O'Connell JE, Thomson RG (2003) A systematic review of cognitive decline in the general elderly population. Int J Geriatr Psychiatry 18:1121-1134.

Passow S, Westerhausen R, Wartenburger I, Hugdahl K, Heekeren HR, Lindenberger U, Li SC (2012) Human aging compromises attentional control of auditory perception. Psychol Aging 27:99-105.

Passow S, Westerhausen R, Hugdahl K, Wartenburger I, Heekeren HR, Lindenberger U, Li SC (2014) Electrophysiological correlates of adult age differences in attentional control of auditory processing. Cereb Cortex 24:249-260.

Passow S, Muller M, Westerhausen R, Hugdahl K, Wartenburger I, Heekeren HR, Lindenberger U, Li SC (2013) Development of attentional control of verbal auditory perception from middle to late childhood: comparisons to healthy aging. Dev Psychol 49:1982-1993.

Peelle JE, Troiani V, Grossman M, Wingfield A (2011) Hearing loss in older adults affects neural systems supporting speech comprehension. J Neurosci 31:12638-12643.

Pessoa L, Kastner S, Ungerleider LG (2003) Neuroimaging studies of attention: from modulation of sensory processing to top-down control. J Neurosci 23:3990-3998.

Petersen EB, Wöstmann M, Obleser J, Stenfelt S, Lunner T (2015) Hearing loss impacts neural alpha oscillations under adverse listening conditions. Front Psychol 6:177.

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. Nat Neurosci 7:658-663.

Pichora-Fuller MK (2003a) Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. Int J Audiol 42 Suppl 1:S59-67.

Pichora-Fuller MK (2003b) Cognitive aging and auditory information processing. Int J Audiol 42 Suppl 2:2S26-32.

Pichora-Fuller MK, Souza PE (2003) Effects of aging on auditory processing of speech. Int J Audiol 42 Suppl 2:2S11-16.

Pichora-Fuller MK, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97:593-608.

Picou EM, Ricketts TA (2014) Increasing motivation changes subjective reports of listening effort and choice of coping strategy. Int J Audiol 53:418-426.

Picton TW, Hillyard SA (1974) Human auditory evoked potentials. II. Effects of attention. Electroencephalogr Clin Neurophysiol 36:191-199.

Polk TA, Drake RM, Jonides JJ, Smith MR, Smith EE (2008) Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: a functional magnetic resonance imaging study of the Stroop task. J Neurosci 28:13786-13792.

Pourtois G, Schwartz S, Seghier ML, Lazeyras F, Vuilleumier P (2006) Neural systems for orienting attention to the location of threat signals: an event-related fMRI study. Neuroimage 31:920-933.

Ptak R (2012) The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. Neuroscientist 18:502-515.

Raichle ME (2011) The restless brain. Brain Connect 1:3-12.

Rebert CS, McAdam DW, Knott JR, Irwin DA (1967) Slow potential change in human brain related to level of motivation. J Comp Physiol Psychol 63:20-23.

Reuter-Lorenz PA, Cappell KA (2008) Neurocognitive Aging and the Compensation Hypothesis. Current Directions in Psychological Science 17:177-182.

Rockstroh B, Muller M, Wagner M, Cohen R, Elbert T (1993) "Probing" the nature of the CNV. Electroencephalogr Clin Neurophysiol 87:235-241.

Rohenkohl G, Nobre AC (2011) alpha oscillations related to anticipatory attention follow temporal expectations. J Neurosci 31:14076-14084.

Rönnberg J, Lunner T, Zekveld A, Sorqvist P, Danielsson H, Lyxell B, Dahlstrom O, Signoret C, Stenfelt S, Pichora-Fuller MK, Rudner M (2013) The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Front Syst Neurosci 7:31.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367-373.

Rosenthal R (1994) Parametric measures of effect size. In: Handbook of research synthesis (Cooper H, Hedges LV, eds), pp 231-244. New York: Russell Sage Foundation.

Rosenzweig MR (1951) Representations of the two ears at the auditory cortex. Am J Physiol 167:147-158.

Rotermund D, Taylor K, Ernst UA, Kreiter AK, Pawelzik KR (2009) Attention improves object representation in visual cortical field potentials. J Neurosci 29:10120-10130.

Roux F, Uhlhaas PJ (2014) Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? Trends Cogn Sci 18:16-25.

Rudner M, Lunner T (2014) Cognitive spare capacity and speech communication: a narrative overview. Biomed Res Int 2014:869726.

Ruggles D, Bharadwaj H, Shinn-Cunningham BG (2012) Why middle-aged listeners have trouble hearing in everyday settings. Curr Biol 22:1417-1422.

Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud AL, Kleinschmidt A (2010) Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. J Neurosci 30:10243-10250.

Salame P, Baddeley A (1987) Noise, unattended speech and short-term memory. Ergonomics 30:1185-1194.

Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. Nature Reviews Neuroscience 2:539-550.

Salthouse TA, Kersten AW (1993) Decomposing adult age differences in symbol arithmetic. Mem Cognit 21:699-710.

Sander MC, Werkle-Bergner M, Lindenberger U (2012) Amplitude modulations and inter-trial phase stability of alpha-oscillations differentially reflect working memory constraints across the lifespan. Neuroimage 59:646-654.

Sarter M, Gehring WJ, Kozak R (2006) More attention must be paid: the neurobiology of attentional effort. Brain Res Rev 51:145-160.

Sato A, Yasuda A, Ohira H, Miyawaki K, Nishikawa M, Kumano H, Kuboki T (2005) Effects of value and reward magnitude on feedback negativity and P300. Neuroreport 16:407-411.

Sauseng P, Klimesch W, Heise KF, Gruber WR, Holz E, Karim AA, Glennon M, Gerloff C, Birbaumer N, Hummel FC (2009) Brain oscillatory substrates of visual short-term memory capacity. Curr Biol 19:1846-1852.

Scharinger M, Herrmann B, Nierhaus T, Obleser J (2014) Simultaneous EEG-fMRI brain signatures of auditory cue utilization. Front Neurosci 8:137.

Scheeringa R, Petersson KM, Oostenveld R, Norris DG, Hagoort P, Bastiaansen MC (2009) Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. Neuroimage 44:1224-1238.

Scheibe C, Ullsperger M, Sommer W, Heekeren HR (2010) Effects of parametrical and trial-to-trial variation in prior probability processing revealed by simultaneous electroencephalogram/functional magnetic resonance imaging. J Neurosci 30:16709-16717.

Schlittmeier SJ, Weisz N, Bertrand O (2011) What characterizes changing-state speech in affecting short-term memory? An EEG study on the irrelevant sound effect. Psychophysiology 48:1669-1680.

Schneider BA, Li L, Daneman M (2007) How competing speech interferes with speech comprehension in everyday listening situations. J Am Acad Audiol 18:559-572.

Schneider BA, Daneman M, Murphy DR, See SK (2000) Listening to discourse in distracting settings: the effects of aging. Psychol Aging 15:110-125.

Scholl BJ (2001) Objects and attention: the state of the art. Cognition 80:1-46.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9-18.

Schröger E, Marzecova A, SanMiguel I (2015) Attention and prediction in human audition: a lesson from cognitive psychophysiology. Eur J Neurosci 41:641-664.

Schultz W (2000) Multiple reward signals in the brain. Nat Rev Neurosci 1:199-207.

Schvartz KC, Chatterjee M, Gordon-Salant S (2008) Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. J Acoust Soc Am 124:3972-3988.

Sebastian M, Reales JM, Ballesteros S (2011) Ageing affects event-related potentials and brain oscillations: a behavioral and electrophysiological study using a haptic recognition memory task. Neuropsychologia 49:3967-3980.

Serences JT, Kastner S (2014) A multi-level account of selective attention. In: The Oxford Handbook of Attention (Nobre AC, Kastner S, eds). New York: Oxford University Press.

Seth AK, Barrett AB, Barnett L (2015) Granger causality analysis in Neuroscience and Neuroimaging. Journal of Neuroscience 35:3293-3297.

Shallice T, Stuss DT, Alexander MP, Picton TW, Derkzen D (2008) The multiple dimensions of sustained attention. Cortex 44:794-805.

Shamma S (2001) On the role of space and time in auditory processing. Trends Cogn Sci 5:340-348.

Shamma S, Lorenzi C (2013) On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. J Acoust Soc Am 133:2818-2833.

Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114-123.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-304.

Shea N, Boldt A, Bang D, Yeung N, Heyes C, Frith CD (2014) Supra-personal cognitive control and metacognition. Trends Cogn Sci 18:186-193.

Sheldon S, Pichora-Fuller MK, Schneider BA (2008) Effect of age, presentation method, and learning on identification of noise-vocoded words. J Acoust Soc Am 123:476-488.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182-186.

Shinn-Cunningham BG, Best V (2008) Selective attention in normal and impaired hearing. Trends Amplif 12:283-299.

Shomstein S, Yantis S (2006) Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. J Neurosci 26:435-439.

Simon HA (1974) How Big Is a Chunk?: By combining data from several experiments, a basic human memory unit can be identified and measured. Science 183:482-488.

Slaney M (1993) An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. In. Cupertino, CA.

Small DM, Gitelman D, Simmons K, Bloise SM, Parrish T, Mesulam MM (2005) Monetary incentives enhance processing in brain regions mediating top-down control of attention. Cereb Cortex 15:1855-1865.

Smith DT, Jackson SR, Rorden C (2009) Repetitive transcranial magnetic stimulation over frontal eye fields disrupts visually cued auditory attention. Brain Stimul 2:81-87.

Smith DV, Davis B, Niu K, Healy EW, Bonilha L, Fridriksson J, Morgan PS, Rorden C (2010) Spatial attention evokes similar activation patterns for visual and auditory stimuli. J Cogn Neurosci 22:347-361.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87-90.

Snyder AC, Foxe JJ (2010) Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. J Neurosci 30:4024-4032.

Sondergaard P, Majdak P (2013) The Auditory Modeling Toolbox. In: The Technology of Binaural Listening (Blauert J, ed), pp 33-56. Berlin, Heildelberg: Springer.

Spaak E, Bonnefond M, Maier A, Leopold DA, Jensen O (2012) Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. Curr Biol 22:2313-2318.

Spitzer B, Fleck S, Blankenburg F (2014) Parametric alpha- and beta-band signatures of supramodal numerosity information in human working memory. J Neurosci 34:4293-4302.

Spitzer H, Desimone R, Moran J (1988) Increased attention enhances both behavioral and neuronal performance. Science 240:338-340.

Stone MA, Füllgrabe C, Moore BC (2012) Notionally steady background noise acts primarily as a modulation masker of speech. J Acoust Soc Am 132:317-326.

Strauß A, Wöstmann M, Obleser J (2014) Cortical alpha oscillations as a tool for auditory selective inhibition. Frontiers in Human Neuroscience 8:350.

Strauß A, Henry MJ, Scharinger M, Obleser J (2015) Alpha phase determines successful lexical decision in noise. J Neurosci 35:3256-3262.

Studebaker GA (1985) A "rationalized" arcsine transform. J Speech Hear Res 28:455-462.

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403-409.

Syka J (2002) Plastic changes in the central auditory system after hearing loss, restoration of function, and during learning. Physiol Rev 82:601-636.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3:151-162.

Taulu S, Kajola M, Simola J (2004) Suppression of interference and artifacts by the Signal Space Separation Method. Brain Topogr 16:269-275.

Tecce JJ, Scheff NM (1969) Attention reduction and suppressed direct-current potentials in the human brain. Science 164:331-333.

Tecce JJ, Savignano-Bowman J, Meinbresse D (1976) Contingent negative variation and the distraction--arousal hypothesis. Electroencephalogr Clin Neurophysiol 41:277-286.

Tervaniemi M, Hugdahl K (2003) Lateralization of auditory-cortex functions. Brain Res Brain Res Rev 43:231-246.

Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J Neurosci 26:9494-9502.

Travis F, Tecce JJ (1998) Effects of distracting stimuli on CNV amplitude and reaction time. Int J Psychophysiol 31:45-50.

Treisman AM (1960) Contextual Cues in Selective Listening. Quarterly Journal of Experimental Psychology 12:242-248.

Treisman AM (1964) Monitoring and Storage of Irrelevant Messages in Selective Attention. Journal of Verbal Learning and Verbal Behavior 3:449-459.

Tremblay K, Ross B (2007) Effects of age and age-related hearing loss on the brain. J Commun Disord 40:305-312.

Tremblay KL, Piskosz M, Souza P (2003) Effects of age and age-related hearing loss on the neural representation of speech cues. Clin Neurophysiol 114:1332-1343.

Tremblay S, MacKen WJ, Jones DM (2001) The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. Memory 9:323-331.

Tremblay S, Nicholls AP, Alford D, Jones DM (2000) The irrelevant sound effect: does speech play a special role? J Exp Psychol Learn Mem Cogn 26:1750-1754.

Treue S (2001) Neural correlates of attention in primate visual cortex. Trends Neurosci 24:295-300.

Treue S (2003) Visual attention: the where, what, how and why of saliency. Curr Opin Neurobiol 13:428-432.

Tun PA, O'Kane G, Wingfield A (2002) Distraction by competing speech in young and older adult listeners. Psychol Aging 17:453-467.

Tun PA, McCoy S, Wingfield A (2009) Aging, hearing acuity, and the attentional costs of effortful listening. Psychol Aging 24:761-766.

Vallar G (1998) Spatial hemineglect in humans. Trends Cogn Sci 2:87-97.

Vallar G, Perani D (1987) The Anatomy of Spatial Neglect in Humans. Advances in Psychology 45:235-258.

Van Boxtel GJM, Böcker KBE (2004) Cortical Measures of Anticipation. Journal of Psychophysiology 18:61-76.

van Dijk H, Schoffelen JM, Oostenveld R, Jensen O (2008) Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. J Neurosci 28:1816-1823.

van Ede F, de Lange F, Jensen O, Maris E (2011) Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha- and beta-band oscillations. J Neurosci 31:2016-2024.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis MA, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci U S A 111:14332-14341.

Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867-880.

Vogel EK, Luck SJ, Shapiro KL (1998) Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. J Exp Psychol Hum Percept Perform 24:1656-1674.

Vogel EK, McCollough AW, Machizawa MG (2005) Neural measures reveal individual differences in controlling access to working memory. Nature 438:500-503.

Vrba J, Robinson SE (2001) Signal processing in magnetoencephalography. Methods 25:249-271.

Walter WG, Cooper R, Aldridge VJ, McCallum WC, Winter AL (1964) Contingent Negative Variation: An Electric Sign of Sensorimotor Association and Expectancy in the Human Brain. Nature 203:380-384.

Walz JM, Goldman RI, Carapezza M, Muraskin J, Brown TR, Sajda P (2013) Simultaneous EEG-fMRI reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. J Neurosci 33:19212-19222.

Wechsler D (1984) Manual for the Wechsler Adult Intelligence Scale-Revised. In. San Antonio, TX: Psychological Corp.

Weinberg H (1972) The contingent negative variation: its relation to feedback and expectant attention. Neuropsychologia 10:299-306.

Weisz N, Müller N, Jatzev S, Bertrand O (2014) Oscillatory alpha modulations in right auditory regions reflect the validity of acoustic cues in an auditory spatial attention task. Cereb Cortex 24:2579-2590.

Weisz N, Hartmann T, Müller N, Lorenz I, Obleser J (2011) Alpha rhythms in audition: cognitive and clinical perspectives. Front Psychol 2:73.

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010-14021.

Wilkinson RT, Ashby SM (1974) Selective attention, contingent negative variation and the evoked potential. Biol Psychol 1:167-179.

Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2014) Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. Cereb Cortex Advance Access, doi:10.1093/cercor/bhu004.

Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2015) Slow-delta phase concentration marks improved temporal expectations based on the passage of time. Psychophysiology, doi: 10.1111/psyp.12413.

Wingfield A, Tun PA, McCoy S (2005) Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. Current directions in Psychological Science 14:144-148.

Winkler I, Teder-Salejarvi WA, Horvath J, Naatanen R, Sussman E (2003) Human auditory cortex tracks task-irrelevant sound sources. Neuroreport 14:2053-2056.

Woldorff M, Hansen JC, Hillyard SA (1987) Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential. Electroencephalogr Clin Neurophysiol Suppl 40:146-154.

Wöstmann M, Schröger E, Obleser J (in press) Acoustic detail guides attention allocation in a selective listening task. Journal of Cognitive Neuroscience, doi: 10.1162/jocn_a_00761.

Wöstmann M, Herrmann B, Wilsch A, Obleser J (2015) Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. J Neurosci 35:1458-1467.

Xu L, Thompson CS, Pfingst BE (2005) Relative contributions of spectral and temporal cues for phoneme recognition. J Acoust Soc Am 117:3255-3267.

Yeshurun Y, Carrasco M (1998) Attention improves or impairs visual performance by enhancing spatial resolution. Nature 396:72-75.

Yeung N, Sanfey AG (2004) Independent coding of reward magnitude and valence in the human brain. J Neurosci 24:6258-6264.

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev 111:931-959.

Yi DJ, Woodman GF, Widders D, Marois R, Chun MM (2004) Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. Nat Neurosci 7:992-996.

Yuan H, Liu T, Szarkowski R, Rios C, Ashe J, He B (2010) Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: an EEG and fMRI study of motor imagery and movements. Neuroimage 49:2596-2606.

Zanto TP, Toy B, Gazzaley A (2010) Delays in neural processing during working memory encoding in normal aging. Neuropsychologia 48:13-25.

Zanto TP, Rubens MT, Thangavel A, Gazzaley A (2011a) Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. Nat Neurosci 14:656-661.

Zanto TP, Pan P, Liu H, Bollinger J, Nobre AC, Gazzaley A (2011b) Age-related changes in orienting attention in time. J Neurosci 31:12461-12470.

Zatorre RJ, Penhune VB (2001) Spatial localization after excision of human auditory cortex. J Neurosci 21:6321-6328.

Zedelius CM, Veling H, Bijleveld E, Aarts H (2012) Promising high monetary rewards for future task performance increases intermediate task performance. PLoS One 7:e42547.

Zion-Golumbic E, Schroeder CE (2012) Attention modulates 'speech-tracking' at a cocktail party. Trends Cogn Sci 16:363-364.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron 77:980-991.

List of figures

Summary

Introduction

A key feature of human environments is that they are rich in sensory information from relevant and irrelevant sources. Human speech is a paradigmatic example of a sensory signal that often occurs in the presence of acoustic interference, resulting from environmental noise or concurrent speakers. There is thus a great research interest in humans' remarkable ability to select relevant speech from the plethora of sounds arriving at the ear (i.e., the "cocktail-party problem", Cherry, 1953). Selective attention describes the mental faculty of selecting currently relevant information for further processing at the expense of distractors (Desimone and Duncan, 1995). Despite the abundance of neuropsychological research on selective attention across sensory modalities, the neural bases of selective attention to speech in noise are not entirely clear. The present work comprises four studies which record listeners' behavioural responses and neural activity using magneto- and electroencephalography (M/EEG) in different speech comprehension tasks. This thesis first identifies complementary neural signatures of selective attention to speech in noise; second, it demonstrates that modulations of neural activity predict listeners' success of speech comprehension; and third, it integrates findings in an initial framework to explain the significance of neural attention mechanisms for speech comprehension under demanding acoustic conditions.

A particularly interesting test case for the neural dynamics of selective attention to speech in noise is the ageing listener. Hearing acuity (Brant and Fozard, 1990) and also cognitive capabilities (Park et al., 2003) show a negative trajectory with age. This might explain listening difficulties of the elderly in complex multi-talker situations (Pichora-Fuller et al., 1995). Although prior behavioural work has shown that speech comprehension in noise decreases at an older age (e.g., Pichora-Fuller and Souza, 2003), it is unresolved in how far these age-differences are explained by changes in older listeners' neural dynamics of selective attention. This thesis compares listeners of different age (younger, 20–30 years; older, 60–80 years) in speech comprehension tasks. In order to contrast neural dynamics between age groups, it is necessary to control for age-differences in sensory hearing acuity which might affect results. Therefore, two studies of this thesis apply careful individual adjustments of acoustic conditions to equalize stimulus audibility and overall task difficulty across age groups. In general, this

thesis asks to what extent neural signatures of attention change at an older age and in how far this explains older listeners' difficulties of speech comprehension in noise.

In three different experimental paradigms, we test the comprehension of spoken digits under acoustic degradation in an overall sample of 98 participants. Acoustic degradation is implemented by concurrent task-irrelevant speech and spectral degradation of the temporal fine structure, an acoustic feature found highly relevant for listening against noise (Hopkins and Moore, 2009, 2010). The high temporal resolution of the M/EEG allows us to study temporally distinct neural mechanisms of attention to speech in noise: We explore stimulus-locked activity in the EEG to investigate listeners' allocation of attention as well as the segregation of attended and unattended speech on a neural level. We analyse brain oscillatory dynamics to investigate age-differences in the use of acoustic and predictive cues to reduce demands on attention. Furthermore, the adequate spatial resolution of MEG allows us to identify brain regions implementing the attentional selection of speech in noise. In the following, I will summarize experiments and results of this thesis.

Experiments and results

The present thesis started out with an electroencephalography (EEG) study of the event-related potential (ERP) in response to degradation of the temporal fine structure in speech materials. Younger (20–30 years) and older participants (60–70 years) performed a numerical comparison of two spoken digits in the presence of task-irrelevant speech. Overall stimulus intensity and the signal-to-noise ratio (SNR) were adjusted to equalize stimulus audibility and overall task difficulty across participants. The results demonstrate that acoustic degradation guides listeners' allocation of attention. In detail, larger amplitudes of the contingent negative variation (CNV) in the EEG indicated that the allocation of attention increased when listeners were faced with more degraded acoustics. This acoustics-driven modulation of the CNV was unchanged in older listeners. This speaks for preserved neural dynamics of attention allocation to speech, at least if overall acoustic conditions are carefully adapted to older listeners' individual requirements.

In a follow-up experiment, we investigated in how far increasing task-relevance affects these neural dynamics of attention allocation in a sample of only younger participants. To this end, correct performance in the auditory number comparison task was rewarded by varying monetary incentives (1, 3, or 5 euro cent). The expectation of higher incentives was thought to enhance task-relevance. Increasing incentives did not affect CNV amplitude or behavioural performance. However, CNV

amplitude was significantly larger in this follow-up experiment compared to the main experiment where no incentives were offered (see above). Larger CNV amplitude thus suggests that the allocation of attention can be increased in a context of high task-relevance.

Subsequently, we analysed brain oscillatory mechanisms of younger and older listeners in the auditory number comparison task. Across age groups, the power of alpha oscillations (~10 Hz) decreased with more acoustic detail but interestingly also when the numerical value of the first digit was better predictive of the second digit. Decreasing alpha power indicates reduced demands on selective attention with more acoustic detail and better predictiveness. Critically, acoustic detail affected behavioural responses and neural alpha power dynamics stronger in the elderly. In agreement with related work (e.g., Passow et al., 2012), this might speak for a decline in attentional control at an older age. Across age groups, alpha power modulations with acoustic detail and predictiveness in the experiment were predictive of listeners' subjectively experienced difficulty in everyday multi-talker situations. This demonstrates that the modulation of neural alpha power explains individual differences in the attentional selection of speech in noise.

In a further analysis of the same dataset, we explored in how far listeners' ongoing EEG signals aligned (i.e., phase-locked) to the acoustic envelopes of the attended spoken digits and the unattended background speech. Younger and older listeners' EEG signals significantly phase-locked to attended and unattended speech, indicating that both speech signals were processed on a neural level. Most importantly, however, substantially different patterns of neural phase-locking were observed for attended and unattended speech. Thus, attention modulates neural responses to acoustic features (i.e., speech envelopes), possibly to segregate task-relevant speech from noise.

In an MEG study, we aimed at investigating *how* neural oscillations support selective attention to speech in noise. In a dichotic listening paradigm, younger participants were attending to four spoken digits on one ear, while four task-irrelevant digits were presented simultaneously on the other ear. Recording sites over the hemisphere ipsilateral to the attended digits showed a relative increase of alpha power, whereas contralateral recording sites showed a relative decrease of alpha power. This alpha power lateralization was thus indicative of listener intent, i.e., a listener's focus of attention to one of two concurrent speech streams. The sources of this alpha power lateralization were found in parietal, frontal, and critically also in auditory cortex regions. Considering the functional inhibition framework of alpha oscillations (Jensen and Mazaheri, 2010), our results suggest that neural activity in supramodal (i.e., parietal) and primary sensory (i.e., auditory cortex) regions is regulated to attain

selective attention to speech in noise. In other words, alpha power modulations implement an *attentional filter mechanism* that enhances the neural processing of spatial and acoustic features of the task-relevant signal and suppresses processing of the noise. Interestingly, alpha power lateralization was not constant during selective attention to ongoing speech but fluctuated at the digit presentation rate (0.67 Hz). Larger amplitudes of these fluctuations predicted better recall of attended digits, suggesting that fluctuating alpha power lateralization at the word rate supports spatial selective attention to speech in noise.

In a final behavioural study, we explored which features of an irrelevant speech signal determine how much it draws attention away from task-relevant speech. In an irrelevant speech paradigm, younger and older participants listened to nine spoken digits which they had to maintain in memory in a retention period. During retention, participants were presented with task-irrelevant speech that was manipulated along two orthogonal dimensions: First, we degraded the acoustic detail (using noise-vocoding) and second, we manipulated final-word predictability (high vs low). We found that more acoustic detail but not higher predictability of irrelevant speech impaired the serial recall of digits after the retention period. Although overall performance decreased in older listeners, the relative increase of distraction from task-irrelevant speech with more acoustic detail was unchanged at an older age. These results indicate that the degree to which task-irrelevant speech draws attention away from task-relevant speech increases with higher intelligibility through more preserved acoustic detail. In contrast, predictable irrelevant speech does not capture attention more than unpredictable irrelevant speech.

Discussion

Our findings demonstrate that selective attention to speech in noise is not implemented by a single brain process but rather by temporally and functionally complementary neural dynamics: The CNV component reflects listeners' allocation of attention in anticipation of a demanding listening task. Differential phase-locking of neural activity to the temporal envelopes of attended and unattended speech constitutes a possible mechanism to segregate a task-relevant acoustic signal from noise. The power of alpha oscillations indicates the inhibition of neural processing of task-irrelevant speech features (i.e., spatial location) in order to reduce noise-interference.

Importantly, neural dynamics are not only modulated by our experimental conditions but they are furthermore indicative of how successfully listeners accomplish the attentional selection of speech in noise. This way, larger amplitude of the CNV, stronger modulation of alpha power with

acoustic/predictive cues, and more pronounced alignment of lateralized alpha power with the word rate all predict improved performance in our speech comprehension tasks. Taken together, our results thus support an initial framework stating that speech comprehension in demanding listening situations depends on whether neural attention mechanisms can compensate for the negative effects of acoustic degradation.

This framework also explains why especially older listeners experience difficulties in multi-talker situations. Adult aging is typically accompanied by a loss of hearing acuity, which increases the degradation of the acoustic input. Thus, older listeners have to upregulate neural attention mechanisms to ensure speech comprehension in noise. But since also attention mechanisms decline at an older age, neural compensation cannot be further enhanced and eventually breaks down. Indeed, this interpretation receives initial support from a recent collaborative study where we could show that older listeners' neural alpha power increases for listeners with mild hearing loss but declines again (i.e., it breaks down) for listeners with moderate hearing loss in an effortful listening task (Petersen et al., 2015).

From a different perspective, this thesis has also implications for interventions to support listeners' speech comprehension in noise. In general, restoration of degraded acoustic input (e.g., with hearing aids) should reduce the need for compensation via neural attention mechanisms. Spare attention capacity could then be used for other currently relevant mental operations. Most importantly, in line with other research (Lunner et al., 2009) our findings imply that interventions to restore impaired hearing could benefit from taking into account neural mechanisms of the user. Future hearing aids could assess neural activity via recordings of electrophysiological data from a few electrodes connected to the hearing aid. If electrophysiological recordings indicate high listening effort (e.g., high alpha power) or a shift of the users' focus of attention to one side (e.g., strong alpha power lateralization), hearing aids could increase the automatic noise-cancellation or adjust the directional microphone to amplify the attended speech signal, respectively. This would offer a way to dynamically adjust the acoustic input according to the neural dynamics of the listener in an individualized way in order to support speech comprehension in complex listening situations.

Zusammenfassung

Einleitung

Unsere Umwelt ist reich an sensorischen Informationen, welche von relevanten aber auch irrelevanten Quellen stammen. Die menschliche Sprache ist häufig maskiert durch akustische Störsignale wie Umgebungslärm oder Gespräche im Hintergrund. Es ist daher von großem wissenschaftlichem Interesse zu verstehen, wie Menschen ein relevantes Sprachsignal trotz erheblicher Störgeräusche verstehen können (häufig beschrieben als das sogenannte "Cocktail-Party Problem"; Cherry, 1953). Selektive Aufmerksamkeit beschreibt die kognitive Fähigkeit, relevante Informationen zu selektieren und irrelevante Informationen zu ignorieren (Desimone and Duncan, 1995). Trotz ausgiebiger Erforschung selektiver Aufmerksamkeitsprozesse in den verschiedenen sensorischen Modalitäten ist die neurale Grundlage der aufmerksamkeitsgesteuerten Selektion von Sprache im Störschall unklar. Diese Arbeit umfasst vier Studien, welche Verhaltensdaten und Hirnaktivität mittels der Magnet-/Elektroenzephalographie (M/EEG) messen, während Probanden verschiedene Sprachverständnisaufgaben lösen. Zunächst identifiziert diese Arbeit neurale Mechanismen der selektiven Aufmerksamkeit beim Sprachverstehen. Darüber hinaus wird gezeigt, dass die Modulation neuraler Aktivität das erfolgreiche Verstehen von Sprache im Störschall vorhersagt. Letztlich werden die Ergebnisse dieser Arbeit in ein Modell integriert, welches die entscheidende Rolle neuraler Aufmerksamkeitsmechanismen für das Sprachverstehen in akustisch anspruchsvollen Situationen erklärt.

Besonders interessant für die Erforschung neuraler Mechanismen des Sprachverstehens im Störschall sind ältere Menschen. Es ist bekannt, dass sowohl die Hörfähigkeit als auch die allgemeine kognitive Leistungsfähigkeit im Alter beeinträchtigt sind (Brant and Fozard, 1990; Park et al., 2003). Dies könnte erklären, warum gerade ältere Menschen oft Sprachverständnisprobleme in komplexen Hörsituationen mit mehreren Sprechern haben (Pichora-Fuller et al., 1995). Verhaltensstudien haben gezeigt, dass das Sprachverständnis im Störschall im Alter beeinträchtigt ist (z.B., Pichora-Fuller and Souza, 2003). Es ist allerdings unklar, inwiefern dieser Alterseffekt durch veränderte neurale Mechanismen bedingt ist. Diese Arbeit vergleicht Probanden unterschiedlichen Alters (jünger: 20–30 Jahre, älter: 60–80 Jahre) bei verschiedenen Sprachverständnisaufgaben. Um neurale Mechanismen zwischen Altersgruppen zu vergleichen, ist es nötig, Altersunterschiede in der Hörfähigkeit zu

kontrollieren, welche ebenfalls die Ergebnisse beeinflussen können. Daher verwenden zwei Studien dieser Arbeit sorgfältige Anpassungen der akustischen Stimuli mit dem Ziel, die Hörbarkeit und die Aufgabenschwierigkeit für Probanden unterschiedlichen Alters anzugleichen. Diese Arbeit erforscht, inwiefern altersbedingte Veränderungen neuraler Aufmerksamkeitsmechanismen die Sprachverständnisprobleme älterer Menschen erklären können.

In drei unterschiedlichen Paradigmen untersuchen wir das Sprachverständnis gesprochener Zahlenwörter unter akustisch reduzierten Bedingungen in einer Gesamtstichprobe von 98 Probanden. Die Reduzierung des akustischen Signals wird implementiert durch Maskierung mit Störschall und die Reduzierung spektraler Anteile im Sprachsignal, welche besonders wichtig für das Sprachverstehen im Störschall sind (Hopkins and Moore, 2009, 2010). Die hohe zeitliche Auflösung des M/EEG erlaubt die Analyse zeitlich getrennter neuraler Mechanismen selektiver Aufmerksamkeit: Wir analysieren Stimulus-gekoppelte neurale Aktivität um die Aufmerksamkeitszuweisung und die neurale Trennung relevanter und irrelevanter Sprachsignale zu untersuchen. Oszillationen im M/EEG werden quantifiziert, um mögliche Altersunterschiede in der Verwendung akustischer Information und der Vorhersagbarkeit der Sprachstimuli zu erforschen. Außerdem ermöglicht die gute räumliche Auflösung des MEG die Bestimmung der Hirnregionen, welche die aufmerksamkeitsgesteuerte Selektion von Sprache implementieren. Im Folgenden werden Experimente und Ergebnisse dieser Arbeit kurz zusammengefasst.

Experimente und Ergebnisse

Die erste Studie dieser Arbeit untersuchte den Effekt reduzierter Akustik (Reduzierung der zeitlichen Feinstruktur im Sprachsignal) auf das ereigniskorrelierte Potential (EKP) im EEG. Jüngere (20–30 Jahre) und ältere Probanden (60–70 Jahre) hatten die Aufgabe, zwei gesprochene Zahlen trotz der Störung durch ein gleichzeitig abgespieltes Hörbuch zu verstehen und numerisch zu vergleichen. Die Gesamtlautstärke der Sprachstimuli und das Signal-zu-Rausch Verhältnis wurden individuell angepasst, um die Hörbarkeit der Stimuli und die Aufgabenschwierigkeit für alle Probanden anzugleichen. Die Ergebnisse zeigten, dass die akustischen Bedingungen die Aufmerksamkeitszuweisung der Probanden beeinflussten. Eine größere Amplitude der *Contingent negative variation* (CNV) im EEG zeigte eine erhöhte Zuweisung selektiver Aufmerksamkeit an, wenn die Zuhörer mit stärker reduzierten akustischen Bedingungen konfrontiert waren. Diese Akustik-induzierte Modulation der CNV war unverändert in der Gruppe der älteren Probanden. Dies spricht

für die Erhaltung grundlegender neuraler Mechanismen der Aufmerksamkeitszuweisung im Alter, zumindest wenn die akustischen Bedingungen sorgfältig an die individuellen Bedürfnisse der Probanden angepasst sind.

In einer Folgestudie untersuchten wir, inwiefern eine erhöhte Relevanz der Sprachverständnisaufgabe die neuralen Mechanismen der Aufmerksamkeitszuweisung beeinflusst. Eine Gruppe jüngerer Probanden führte dazu den auditiven Zahlenvergleich (siehe oben) durch, wobei der korrekte numerische Vergleich der beiden Zahlen in jedem Durchgang finanziell belohnt wurde (mit 1, 3, oder 5 Eurocent). Die Aussicht auf eine höhere finanzielle Belohnung sollte die Relevanz der Sprachverständnisaufgabe erhöhen. Die finanzielle Belohnung hatte keinen Einfluss auf die CNV Amplitude oder auf die Leistung der Probanden. Allerdings war die CNV Amplitude in dieser Folgestudie insgesamt größer verglichen mit dem Hauptexperiment, in dem keine finanzielle Belohnung dargeboten wurde. Die erhöhte CNV Amplitude in dieser Folgestudie deutet somit darauf hin, dass neurale Mechanismen der Aufmerksamkeitszuweisung in einem Kontext hoher Aufgabenrelevanz verstärkt werden können.

Anschließend analysierten wir neurale Oszillationen im EEG jüngerer und älterer Probanden beim auditiven Zahlenvergleich. In beiden Altersgruppen verringerte sich die Power der Alpha Oszillationen (~10 Hz) wenn die Akustik weniger stark reduziert war, aber auch wenn der numerische Wert der ersten Zahl eine bessere Vorhersage der zweiten Zahl ermöglichte. Verringerte Alpha Power deutet darauf hin, dass die Anforderung an die aufmerksamkeitsgesteuerte Selektion der gesprochenen Zahlen mit besserer Akustik und besserer Vorhersagbarkeit des Sprachsignals abnahm. Entscheidend bei diesen Ergebnissen war, dass sowohl die Leistung als auch die Alpha Power älterer Probanden stärker abhängig von dem Grad der akustischen Reduzierung war. Im Einklang mit anderen Studien (z.B., Passow et al., 2012) sprechen unsere Ergebnisse daher für eine verminderte Kontrolle von Aufmerksamkeitsprozessen im Alter. Unabhängig vom Alter sagten die Alpha Power Modulationen im Experiment voraus, wie schwer den Probanden das Sprachverstehen in alltäglichen Hörsituationen im Störschall fällt. Zusammenfassend demonstriert diese Studie, dass Modulationen der neuralen Alpha Oszillationen individuelle Unterschiede selektiver Aufmerksamkeitsprozesse erklären.

In einer weiteren Analyse derselben Daten wurde untersucht, inwieweit Fluktuationen im EEG Signal die Amplitudenmodulationen des Sprachsignals abbilden. Zu diesem Zweck berechneten wir die Kreuzkorrelation zwischen Amplitudenmodulationen (der sogenannten *Einhüllenden*) der gesprochenen Zahlen und des störenden Hörbuches mit dem EEG Signal. Interessanterweise

korrelierte das EEG Signal sowohl mit der Einhüllenden der Zahlen als auch mit der Einhüllenden des störenden Hörbuches. Dies belegt, dass beachtete aber auch ignorierte Sprachsignale neural verarbeitet werden. Entscheidend jedoch war, dass die Kreuzkorrelation eine negative Korrelation mit der Einhüllenden der beachteten Zahlen und eine positive Korrelation mit der Einhüllenden des ignorierten Hörbuches zeigte. Somit moduliert Aufmerksamkeit die neurale Verarbeitung gleichzeitig präsentierter Sprachsignale, möglicherweise um relevante Sprache vom Störschall zu trennen.

In einer MEG Studie untersuchten wir die Funktion neuraler Oszillationen für die aufmerksamkeitsgesteuerte Selektion von Sprache im Störschall. In einer dichotischen Höraufgabe richteten jüngere Probanden ihre Aufmerksamkeit auf vier gesprochene Zahlen, welche entweder auf dem rechten oder auf dem linken Ohr präsentiert wurden. Gleichzeitig wurden auf dem anderen Ohr vier irrelevante Zahlen präsentiert. Richteten die Probanden ihre Aufmerksamkeit auf die Zahlen auf der linken Seite, so stieg die Alpha Power in der linken Hemisphäre an und nahm in der rechten Hemisphäre ab. Diese Alpha Lateralisierung kehrte sich um, wenn Probanden ihre Aufmerksamkeit auf die Zahlen auf der rechten Seite richteten. Die lateralisierte Power der Alpha Oszillationen zeigte somit an, welches von zwei gleichzeitig präsentierten Sprachsignalen vom Zuhörer beachtet wurde. Die neuralen Quellen der Alpha Lateralisierung umfassten frontale und parietale Regionen, wie auch Regionen des auditiven Cortex. Diese Ergebnisse zeigen, dass neurale Aktivität in supramodalen (parietalen) und auch sensorischen (auditiven) Hirnregionen moduliert wird um die aufmerksamkeitsgesteuerte Selektion von Sprache im Störschall zu ermöglichen. Es ist anzumerken, dass die Alpha Lateralisierung mit der Präsentationsrate der Zahlen fluktuierte (0.67 Hz). Waren diese Fluktuationen stärker ausgeprägt, so verbesserte sich auch die korrekte Wiedergabe der beachteten Zahlen durch die Probanden.

Letztlich führten wir eine Verhaltensstudie durch, um zu testen, welche Eigenschaften eines störenden Sprachsignals beeinflussen, wie gut dieses von jüngeren und älteren Probanden ignoriert werden kann. Die Probanden hörten neun gesprochene Zahlen, welche sie sich in der präsentierten Reihenfolge merken sollten. Anschließend waren drei störende Sätze zu hören, bevor die Probanden die Zahlen in der Reihenfolge ihrer Präsentation auswählen mussten. Das störende Sprachsignal wurde orthogonal in zwei Dimensionen verändert: Erstens reduzierten wir die Akustik durch Verminderung der spektralen Information (durch *noise-vocoding*). Zweitens war die Vorhersagbarkeit des letzten Wortes in den störenden Sätzen entweder hoch oder gering. Die Ergebnisse zeigten, dass eine bessere Akustik des störenden Sprachsignals, nicht aber eine höhere Vorhersagbarkeit des störenden

Sprachsignals, die Erinnerung an die korrekte Reihenfolge der Zahlen negativ beeinflusste. Trotz insgesamt schlechterer Leistungen war dieses Verhaltensmuster bei älteren Probanden unverändert. Unsere Ergebnisse sprechen dafür, dass ein störendes Sprachsignal die Aufmerksamkeit für relevante Sprache besonders stark einschränkt, wenn es von akustisch besserer Qualität und damit besser verständlich ist.

Diskussion

Die Ergebnisse dieser Arbeit zeigen, dass die aufmerksamkeitsgesteuerte Selektion von Sprache im Störschall nicht durch einen einzelnen, sondern durch mehrere zeitlich und funktionell unterschiedliche neurale Mechanismen realisiert wird: Die CNV zeigt die Zuweisung der Aufmerksamkeit in Erwartung einer schwierigen Höraufgabe an. Entgegengesetzte Korrelationen des EEG Signals mit der Einhüllenden von beachteter und ignorierter Sprache stellen einen möglichen Mechanismus zur neuralen Trennung relevanter Sprache vom Störschall dar. Die Power der Alpha Oszillationen reflektiert die Inhibition der neuralen Verarbeitung von Eigenschaften eines störenden Sprachsignals (z.B., räumliche Position), um das Sprachverstehen relevanter Sprache zu unterstützen.

Es ist zu beachten, dass neurale Aktivität nicht nur durch unsere experimentellen Konditionen beeinflusst wurde. Darüber hinaus erklärt die Stärke der neuralen Aktivitätsveränderung den Erfolg der Probanden beim Verstehen von Sprache im Störschall. Größere CNV Amplitude, stärkere Alpha Power Modulation mit besserer Akustik/Vorhersagbarkeit des Sprachsignals und erhöhte Fluktuationen der Alpha Lateralisierung mit der Präsentationsrate der Sprache korrelierten alle mit besserer Leistung in unseren Sprachverständnisaufgaben. Unsere Ergebnisse lassen sich daher in einem Modell vereinen, welches besagt, dass das Sprachverständnis im Störschall davon abhängt, inwiefern neurale Aufmerksamkeitsmechanismen die negativen Effekte akustischer Reduzierung kompensieren.

Dieses Modell erklärt auch, warum besonders ältere Menschen Schwierigkeiten in Hörsituationen mit mehreren Sprechern haben. Im Alter setzt häufig eine Verringerung der Hörfähigkeit ein, welche dazu führt, dass akustische Signale reduziert wahrgenommen werden. Somit müssen ältere Menschen ihre Aufmerksamkeitsmechanismen verstärken, um die akustische Reduzierung durch erhöhte Aufmerksamkeit zu kompensieren. Jedoch kommt es im Alter auch zu einer Verringerung der Kontrolle von Aufmerksamkeitsprozessen, was dazu führt, dass die Kompensation für reduzierte Akustik nicht aufrechterhalten werden kann und schließlich zusammenbricht. Diese Interpretation

wird unterstützt durch eine kürzlich veröffentlichte kollaborative Studie (Petersen et al., 2015), in der wir zeigen konnten, dass die Alpha Power bei einer anspruchsvollen Höraufgabe mit mäßigem Hörverlust ansteigt, jedoch mit stärkerem Hörverlust zusammenbricht.

Betrachtet man diese Arbeit aus einem anderen Blickwinkel, bietet sie Ansätze für Interventionsmöglichkeiten um das Sprachverstehen im Störschall zu erleichtern. Generell sollte die Wiederherstellung des akustischen Signals (z.B. mithilfe von Hörgeräten) die Anforderungen an neurale Aufmerksamkeitsmechanismen verringern. Im Einklang mit anderen Studien (Lunner et al., 2009) implizieren unsere Ergebnisse, dass Interventionen zur Verbesserung des Sprachverstehens im Störschall von der Berücksichtigung neuraler Aufmerksamkeitsmechanismen profitieren können. Zukünftige Hörgeräte könnten neurale Aktivität zum Beispiel mithilfe integrierter EEG Elektroden messen. Zeigen diese Messungen eine erhöhte Höranstrengung an (z.B. große CNV Amplitude oder erhöhte Alpha Power), so könnte das Hörgerät die automatische Rauschunterdrückung verstärken. Zeigen die EEG Messungen an, dass der Zuhörer seine Aufmerksamkeit auf ein Sprachsignal auf einer Seite richtet (z.B. deutliche Alpha Lateralisierung), so könnte das Hörgerät das Richtmikrofon demensprechend einstellen, um das beachtete Signal zu verstärken. Somit wäre es möglich, das akustische Signal individuell und dynamisch an die neuralen Aufmerksamkeitsprozesse des Zuhörers anzupassen, um das Sprachverstehen in komplexen Hörsituationen zu unterstützen.

Curriculum vitae

- **Personal Data**

Name:	Malte Wöstmann
Date of birth:	2nd of April, 1986
Civil status:	unmarried
Children:	–
Nationality:	German

- **Education and professional experience**

Since 04/2015	**Research associate (Post-Doc)** at the University of Lübeck, Institute for Psychology
08/2012-03/2015	**PhD candidate** in the International Max Planck Research School (IMPRS) for Neuroscience of Communication, Leipzig
05/2012-09/2012	**Research associate** at the University of Osnabrück, Institute for Cognitive Science
04/2010-03/2012	**M. Sc. Cognitive Science**, University of Osnabrück, final grade: excellent (A)
10/2006-02/2010	**B. Sc. Cognitive Science,** University of Osnabrück, final grade: excellent (A)
08/1998-06/2005	**St. Angela academic high school**, Osnabrück, Germany

- **Scholarships and awards**

Colin Cherry best poster award. 6th Workshop on Speech in Noise (SpiN): Intelligibility and Quality. *Marseille, France, January 09–10, 2014.*

PhD scholarship, *International Max Planck Research School (IMPRS) for Neuroscience of Communication, 01/2013-03/2015.*

One-year scholarship of the "graduate school adaptivity of hybrid cognitive systems". *University of Osnabrück, 04/2011–03/2012.*

Scholarship for excellent study achievements. *University of Osnabrück, 01/2011.*

- **Publications in peer review journals**

Wöstmann, M., Schröger, E., & Obleser, J. (in press). Acoustic detail guides attention allocation in a selective listening task. *Journal of Cognitive Neuroscience*.

Borch-Petersen, E., **Wöstmann, M.**, Stenfelt, S., & Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Frontiers in Psychology*, 6:177.

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience*, 35 (4): 1458-1467.

Strauß, A.†, **Wöstmann, M.**†, & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8:350. † both authors contributed equally.

Obleser, J., **Wöstmann, M.**, Hellbernd, N., Wilsch, A., Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, 32 (36):12376 – 12383.

- **Public talks**

Wöstmann, M. Neural oscillations reflect attentional challenges of understanding speech in noise. *7th Speech in Noise (SpiN) Workshop, Copenhagen, Denmark, January 8–9, 2015.*

Wöstmann, M., Herrmann, B., Schröger, E., Obleser, J. Cognitive capabilities and listening in demanding conditions. *Siemens Symposium, Erlangen, Germany, February 27–28, 2014.*

Wöstmann, M. Gehirnmechanismen selektiven Hörens. *Oticon Symposium, Hamburg, Germany, November 13, 2013.*

Wöstmann, M., Krause, C. D., König, P., Bosch, P. An ERP Study of Linguistic Preference and World Knowledge Plausibility. *Experimental Pragmatics Conference, Barcelona, Spain, June 04, 2011.*

Selbständigkeitserklärung gemäß § 8(2) der Promotionsordnung

Hiermit versichere ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde, und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind. Ich versichere, dass die vorliegende Arbeit in gleicher oder in ähnlicher Form keiner anderen wissenschaftlichen Einrichtung zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Es haben keine früheren erfolglosen Promotionsversuche stattgefunden. Die Promotionsordnung der Fakultät für Biowissenschaften, Pharmazie und Psychologie der Universität Leipzig ist mir bekannt und ich erkenne diese an.

Leipzig, den 16.04.2015

_____

Malte Wöstmann